Differentially Private Recommender Systems(差分隐私与推荐系统):

Building Privacy into the Netflix Prize Contenders(在 Netflix 奖金竞争作品中加入隐私保护)

Frank McSherry and Ilya Mironov

MicrosoftResearch, Silicon Valley Campus

{mcsherry, mironov}@microsoft.com

摘要：

我们考虑从集体用户行为中提出建议的问题，同时为这些用户提供隐私保证。具体而言，我们认为 Netflix Prize 数据集及其领先算法适用于差分隐私框架。不像以前的隐私工作涉及密码保护推荐计算，差分隐私限制了计算的方式，排除了对输出中潜在记录的任何推断。这些算法必然引入不确定性 - 即噪声 - 计算，隐私交易的准确性。我们发现，Netflix 奖竞赛中的几种主要方法可以进行调整，以提供差分隐私，而不会显着降低其准确性。为了适应这些算法，我们明确地将它们分为两部分，可以用差分隐私保证执行的聚合/学习阶段，以及使用学习的相关性和个人数据提供个性化推荐的个人推荐阶段。这些改编不是微不足道的，而是涉及仔细分析算法的每记录灵敏度以校准噪声，以及新的后处理步骤以减轻这种噪声的影响。我们测量了这些适应性中准确性和隐私之间的经验交易 - 并发现我们可以提供非平凡的形式隐私保证，同时仍然超越了 Netflix 提供的 Cinematch 基线。

分类与题目描述：

H.2.8 [数据库管理]: [数据挖掘]

通用标签：

算法，安全，理论

关键词：

差分隐私，Netflix，推荐系统

1. 动机：

基于协作过滤的推荐系统是一把双刃剑。通过汇总和处理多个用户的偏好，它可以提供相关的建议，提高网站的收入并增强用户体验。在流量方面，它是用户共享私人信息泄漏的潜在来源。本文的重点是设计，分析和实验验证推荐系统内置隐私保证。我们在 Netflix Prize 数据集上测量系统的准确性，这也推动了我们对算法的选择。提高推荐系统的准确性并为其用户提供隐私的目标很好地一致。它们是良性循环的一部分，更高的准确性和更强的隐私保证可以减轻与分享个人信息相关的焦虑，从而导致更广泛和更深入的参与，从而提高准确性和隐私。考虑收集，存储和处理来自其用户群的信息的推荐系统。即使所有安全措施（如适当的访问控制机制，受保护的存储，加密的客户端 - 服务器通信都已到位），系统对任何用户都可见的输出（即推荐）部分来自其他用户的输入。好奇的或恶意的用户或其联盟可能试图根据他们自己的观点以及通过推荐系统的标准接口暴露的观点来推断别人的输入。这种威胁在开放访问网站的情况下特别不明显，因为身份弱和在线主动攻击的可能性更大，攻击者可以创建多个帐户，将其自己的输入注入推荐系统，观察更改并适应其行为，仅受网络速度和系统周转时间的限制。有两个常见的论点用于反映推荐系统提出的隐私问题。我们依次解决这些争论。

数据不敏感。在许多情况下，用户共享的信息被认为是不敏感的，并被视为如此。我们观察到数据的敏感性是上下文的，很大程度上取决于用户的情况和攻击者的腋窝知识，任何全局策略决策在保守方面都应该是错误的。不成熟的用户可能并不知道在日常网络冲浪过程中（例如 IP 地址，计时信息，HTTP 头等）提供（并且经常收集）的个人数据的数量，这些数据具有深远的隐私影响。此外，正确判断信息的敏感性是困难的，本身就是一个移动的目标。 Narayanan 和 Shmatikov [22]展示了一个强大且实用的反匿名攻击，将 Netflix 奖数据集中的记录与公共 IMDb 配置文件相链接。进一步考虑他们的攻

击，考虑一个情景，其中一个人通过偶尔讨论不相交的电影集合来维护两个不同的配置文件（例如，专业博客和社交网络上的假名页面），所有这些都以 Netflix 评级。那么他们可能会被链接到同一个 Netflix 奖励行，从而彼此链接。无论是归因于这个人的电影评级是尴尬，显露还是显而易见都是无关紧要的，事实上这两个人物角色是相互联系的，可以被认为是深度侵入性和令人不安的。

质量推荐系统的隐含性质。一个对一个人的输入过于敏感的推荐系统通常被认为是不合适的。但是，在评估系统的安全性时，必须考虑不是典型的状态，而是考虑所有可行的状态，一个有决心和足够资源的攻击者可以迫使它进入状态。例如，在基于用户 - 用户相似性的推荐系统中，攻击者可以根据攻击者的部分知识创建一个类似于个人的简单配置文件（或多个配置文件），并诱导推荐系统忠实地报告个人高度评价的所有项目。允许提交条目的系统（例如 digg.com 或 stumbleupon.com）即使只依赖项目项目的相似性，也容易受到类似攻击。尽管这种攻击的复杂性随着推荐系统的复杂性而增加，而推荐系统通常被保密，但通过默默无闻的安全性是一种不可信的安全措施。换句话说，隐私保护推荐系统必须保留其安全属性，以抵御任何可以访问系统设计的可行攻击者（或一组明确定义的攻击者）以及有关其目标的不受限制的辅助信息。

1.1. 研究意义：

这项工作的主要贡献是设计和分析一个现实的推荐系统，以提供现代隐私保证。这项任务并非微不足道：先前推荐系统的设计并非着眼于隐私，事先隐私研究的重点是更适度的算法，而没有尝试实际验证。推荐系统增加了额外的复杂性，即它们的端对端行为应该反映每个用户的私人数据，这要求我们将隐私保护“学习”阶段与高度非私人“预测”阶段（由用户进行，在他们自己家中的隐私）。一种自然的方法将遵循匿名数据发布中发生的大量工作，例如 k-匿名[23]，其中，如 Netflix 所做的那样，数据被释放，企图移除敏感信息和过度特定的属性组合。除了不确定的隐私保证[18]，Brickell 和 Shmatikov [8]发现应用于高维数据的这些技术对数据挖掘算法造成了不可挽回的损失。相反，我们将隐私保护整合到计算本身中，从而确保学习到的建议可以使用差分隐私框架来保护隐私。我们的发现是，隐私不需要花费大量精力。对于我们考虑的方法，隐私保护算法可以被参数化以基本匹配其非私有类似物的推荐性能。虽然需要进行一些专业分析，但方法本身相对简单易行。作为本说明的另一个贡献，我们希望展示现代隐私技术与实际和现实的学习系统的整合。

1.2. 相关工作

我们根据 Netflix 奖的领先解决方案选择算法[4,5,6]。我们调整算法，举例说明两种方法，作为 Netflix 奖竞争者的主要组成部分：因子模型和邻域模型。最近几篇论文介绍并研究了 Dwork [14]调查的差分隐私在学习和数据挖掘问题中的应用。算法如 k-均值聚类，感知器分类，关联规则挖掘，决策树归纳和低秩近似都显示出具有不同的私人类似物，其准确性有理论界限[7]。尽管我们从这些作品中大量借用了我们的基础隐私技术，但我们的重点更多的是构建和评估完整的端到端推荐系统，而不是孤立的组件。 Ne fl ix 以匿名用户身份大规模发布数据已被证明具有深远的隐私影响[22]，尤其是确定大多数行可以基于仅仅十几个部分已知的数据点的几乎确定性来识别。尽管一个商业推荐系统不太可能愿意公开其全部或大部分基础数据，但最近 Calandrino 等人的工作。 [9]表明，被动观察亚马逊网站的建议足以对个人的购买历史进行有效的推断。以前关于安全推荐系统问题的密码解决方案的工作重点在于移除单个可信任方访问所有人的数据[10,11,2,3]。它不会试图限制在系统正常执行过程中通过系统建议泄露的信息量。我们的解决方案可以与 Alambic 框架的模块化方法相结合[2]。区分我们的隐私保护计算方法与以前关于隐私研究匿名记录发布的工作很重要。人们可以想象在匿名数据的基础上构建推荐系统或任何机器学习技术，从匿名化中绘制隐私属性，而不是自己再现它们。但是，特别是对于丰富的高维数据，大多数匿名技术似乎会削弱数据的实用性[8,1]。通过将隐私保证集成到应用程序中，我们可以为其提供对原始数据的无限制访问，条件是其最终输出 - 实质上少于整个数据集的信息 - 尊重隐私标准。

2. 推荐算法

我们首先介绍一些适用于 Netflix 奖励的方法。我们所考虑的方法在某一点上是真正的竞争者，但是比现有技术更简单可以理解。虽然他们的准确性水平已被超越，但我们希望通过了解他们的私人调整，

我们可以推导出可能会继续应用于日益复杂的推荐系统的方法。我们考虑的设置包括用户和项目，以及（用户，项目）子集的子集的评分。考虑到这种部分评级，目标是预测特定（用户，物品）位置的某些保留值。

全局效应。这些系统中的第一步是通过计算中心评分并减去用户和项目的平均评分。为了稳定这种计算，通常计算平均值，包括在全局平均数额外的额外数量的评级;用户和电影有许多收视率都会偏移到正确的平均值，但平均值和小的支持不允许过多。

协方差。考虑到评分本身的属性导致的一阶效应，通常需要考虑项目之间的相关性.1 一种常用的方法是查看项目的协方差矩阵，其(i, j)项是平均值所有用户的项目 i 和 j 的评分产品。当然，相对较少的用户实际上对 i 和 j 都进行了评分，因此平均值仅针对评价这两个项目的用户。

几何推荐算法。过度简化，一旦我们计算了项目的协方差矩阵，我们就有足够的信息来应用大量的高级学习和预测算法。协方差矩阵对项目的完整几何描述进行编码，任何几何算法（例如：潜在因子分析，最近邻近方法，几何聚类等）都可以在此处进行部署。重要的是，从我们的角度来看，这些方法可以应用于每个用户，只使用协方差信息和用户收集的评分。如果协方差测量可以私下进行，那么在隐私保证的情况下，可以部署任何不需要返回其他用户的原始数据的算法。我们将尝试几种方法，几乎完全借鉴了之前发表的有关 Netflix 奖的研究成果，但现在暂缓讨论特定算法。

2.1. 非隐私保护解决方案

我们将把前面的草图形式化为一种非私有的算法，但将构成我们隐私保护方法的骨架。 算法中的步骤可能显得格外迂腐，但是以简单的形式编写它们将使我们能够轻松地将它们调整为它们的私有形式。在[4]之后，我们使用 r 来表示评级的集合，其中用于电影 i 的用户 u 的评级的符号r_{u_i}以及用户 u 的评级矢量的r_u。 我们使用e_{u_i}和e_u表示评分的存在（允许我们与报告的零值区分）的二进制元素和向量。我们首先从每部电影中减去电影平均值，其中平均值受到全局平均值的β值的影响。

影片影响

1. 对于每个项目 i，计算总数和计数：

$$\begin{aligned}&\text{(a) Let } \mathrm{MSum}_i = \textstyle\sum_u r_{ui}.\\ &\text{(b) Let } \mathrm{MCnt}_i = \textstyle\sum_u e_{ui}.\end{aligned}$$

2. 计算平均平均值：

$$G = \textstyle\sum_i \mathrm{MSum}_i / \sum_i \mathrm{MCnt}_i.$$

3. 对于每个项目 i，计算标准化平均值：

$$\text{(a) Let } \mathrm{MAvg}_i = (\mathrm{MSum}_i + \beta G)/(\mathrm{MCnt}_i + \beta).$$

4. 对于每个评级 rui，减去适当的平均值：

$$\text{(a) Set } r_{ui} = r_{ui} - \mathrm{MAvg}_i.$$

我们为用户执行完全相同的操作，计算稳定的平均值并从每个评级中减去适当的平均值。 我们的协方差计算也是直接的，但出于明显的原因，我们希望对每个用户的贡献进行加权组合，对于用户 u 使用权重 $0 \leqslant w_u \leqslant 1$：

计算协方差

对于每个电影电影对（i，j）：

$$\begin{aligned}&\text{(a) Let } \mathrm{Cov}_{ij} = \textstyle\sum_u w_u r_{ui} r_{uj}.\\ &\text{(b) Let } \mathrm{Wgt}_{ij} = \textstyle\sum_u w_u e_{ui} e_{uj}.\\ &\text{(c) Let } \mathrm{Avg}_{ij} = \mathrm{Cov}_{ij}/\mathrm{Wgt}_{ij}.\end{aligned}$$

矩阵平均值现在包含我们对协方差矩阵的估计值。然后，我们可以将这个矩阵传递给其他研究人员提出的几种几何方法之一，因为它们对网络数据集尤为有效。 虽然后续算法的选择对于推荐系统的性能显然非常重要，但我们不会尝试从它们的具体规范中推导任何隐私属性。 相反，我们只向他们提供使用差分隐私产生的输入。

3. 差分隐私

在[15]中调查的不完全隐私[13]是基于以下原则的较新的隐私定义：计算输出不应该允许推断计算输入中是否存在任何记录。 在形式上，它要求对于随机计算的任何结果，在有或没有任何一条记录的情况下，结果应该几乎相同。 我们说两个数据集 A 和 B 是相邻的，写成 A≈B，如果一个记录中只有一个记录，而另一个记录不存在。

定义 1：对一函数 M，任何相邻数据集 A 和 B 以及 M 作用于数据集后的可能输出结果的任何子集 S，若有：

$$\mathbf{Pr}[M(A) \in S] \quad \leq \quad \exp(\epsilon) \times \mathbf{Pr}[M(B) \in S] \ .$$

*则称其满足**ε-差分隐私**。*

对保证差异性隐私的一种解释是，它限制了从任何输出事件 S 推断出计算输入是 A 还是 B 的能力。从任意先验 p（A）和 p（B）中，我们可以看到：

$$\frac{p(A|S)}{p(B|S)} = \frac{p(A)}{p(B)} \times \frac{p(S|A)}{p(S|B)} .$$

当 A≈B 时，差分隐私将 exp(ε)的因子更新为前一个因子，从而限制输入数据集中的轻微偏差的推理程度。 具体而言，关于任何单一记录的存在与否（以及因此的价值）的推断都受 exp(ε)的因素限制。

我们强调，差异性隐私是产生输出的计算的一个特性，而不是输出本身的特性。同时，概率纯粹是计算随机性的函数，而不是输入中可能的随机性或不确定性。有大量关于隐私的文献，还有许多其他的定义和方法可以提供其他的保证。其中很多（如流行的 k-匿名）仅为输出提供语法保证，而没有上面使用的语义含义。与大多数其他方法不同，差异性隐私已被证明具有抵御攻击的能力，继续为任意先前知识提供隐私保证，在重复使用情况下针对任意数据类型提供保护。有几种方法已经考虑到差异性隐私的弱化版本，为了提高准确性或可用性，交换了保证的一般性（可能只保护一部分先验者）[21]。尽管如此，我们在这里概述的技术可以提供更强有力的保证，但不清楚是否需要削弱的定义（尽管我们将考虑下一步的放宽）。

3.1. 近似差分隐私

我们还将考虑放宽形式的差异性隐私，允许在[16]中引入一个加性术语，以及乘法术语。

定义 2：如果对于任何相邻的数据集 A 和 B 以及可能结果范围（M）的任何子集 S，随机化计算 M 满足 (ε, δ)不完全隐私。

$$\mathbf{Pr}[M(A) \in S] \leq \exp(\epsilon) \times \mathbf{Pr}[M(B) \in S] + \delta .$$

这种保证的一个解释是，计算 M 的结果不可能提供比ε-差异性隐私更多的信息，但这是可能的。 对于任何γ>ε，取 Sγ作为其结果 x 的集合：

$$\frac{p(x|A)}{p(x|B)} > \exp(\gamma).$$

将这个约束与(ε, δ) - 二阶隐私的定义结合起来，我们可以得出这样的结论是不太可能的：

$$p(S_\gamma|B) \leq p(S_\gamma|A) \leq \frac{\delta}{1-\exp(\epsilon-\gamma)}.$$

尽管γ可能比ε大得多，但概率有效地受δ约束。 此外，对于我们使用的隐私机制（在下一节中介绍），ε和δ之间总是存在交易关系; 一个人可以任意减少，但增加另一个的代价。 从某种意义上说，释放的信息量（以两个概率的比率衡量）是一个最可能很小的随机变量。 重要的是，近似差分隐私满足顺序组合逻辑：

定理 1.如果M_f 和M_g 分别满足$(\varepsilon_f, \delta_f)$和$(\varepsilon_g, \delta_g)$差分隐私，那么它们的线性组合满足$(\varepsilon_f + \varepsilon_g, \delta_f + \delta_g)$−差分隐私。

我们将使用这个定理来推导出我们的推荐系统的端到端隐私保证的界限，它由多个独立的(ε, δ)-差分隐私计算组成。

3.2. 噪音和敏感度

计算数值测量时最简单的差分隐私方法是将随机噪声应用于测量，并认为这掩盖了单个记录对结果的可能影响。 如果我们的目标是计算一个函数 f：$D^n \rightarrow R^d$，下面的结果描述了通过增加噪声实现的先前隐私结果[17]。

定理 2.将 M(X)定义为 f(X)+ Laplace$(0,\sigma)^d$。若：

$$\sigma \geq \max_{A \approx B} \|f(A) - f(B)\|_1 / \epsilon .$$

则 M 在任何时候都可以提供ε-差分隐私。

我们可以使用高斯噪声来实现近似的差分隐私保证，与 f(A)和 f(B)之间的较小||·|| 2 距离成正比。 用平均值μ和方差σ2 写出正态分布的 N（μ，σ2）。 在[16]中证明了这一点。

定理 3.将 M（X）定义为 f(X)+ N$(0,\sigma^2)^d$。当：

$$\sigma \geq \sqrt{2\ln(2/\delta)}/\epsilon \times \max_{A \approx B} \|f(A) - f(B)\|_2.$$

M 提供了(ε, δ)-差分隐私。

注意，对于噪声分布的任何一个参数σ，ε和δ都有很多有效的设置。 具体而言，对于任何正ε，都有一个

与其相关的δ=δ（ε）。 因此，我们通常只会关注σ值，而不会导出特定的（ε，δ）对。

3.3. 计数，均值和协方差

我们需要从数据中测量相对较少的统计数据，以开始调整先前工作中的推荐算法。 全局影响，如每部电影的平均值和每位用户的平均值在预测中发挥重要作用。 此外，电影电影协方差矩阵构成了许多几何算法的基础，特别是 SVD 分解方法和 kNN 几何距离方法。 在继续我们的方法的具体情况之前，我们看到如何在前面描述的框架中测量这些数量。

计数和总和是比较容易分析的函数。 如果 f：$D^n \rightarrow R^d$将记录（评分）划分为 d 个单元并计算每个单元的内容，对于||·|| 1 和||·|| 2：

$$\max_{A \approx B} \|f(A) - f(B)\| \quad = \quad 1 \; .$$

因此，我们可以报告记录的任意分区数（我们的兴趣在于每部电影的评分），并提供适当的附加噪声提供隐私。 总和比较复杂，因为我们必须明确约束每个元素贡献的值的范围。 在收视率的情况下，分数最初为 1 到 5，但随着我们对数据应用各种操作，分数会增加和缩小。 如果单个记录的最大范围至多为 B，那么对于||·|| 1 和||·||2：

$$\max_{A \approx B} \|f(A) - f(B)\| \quad \leq \quad B \; .$$

我们将采用的最复杂的测量是电影电影协方差矩阵。 简化一点（具体地说，现在忽略权重），我们可以将协方差矩阵写为（使用r_u来表示人 u 的评分向量）：

$$\text{Cov} \quad = \quad \sum_u r_u r_u^T \; .$$

这个表达式清楚地表明，单个记录的单个变更对总和的影响可能是有限的。 如果一个评级发生变化，从r_u^a变为r_u^b，两个协方差矩阵之间的差异是：

$$\begin{aligned} \|\text{Cov}^a - \text{Cov}^b\| &= \|r_u^a r_u^{aT} - r_u^b r_u^{bT}\| \\ &\leq \|r_u^a - r_u^b\| \times (\|r_u^a\| + \|r_u^b\|) \; . \end{aligned}$$

设||$r_u^a - r_u^b$||为 1：

$$\|\text{Cov}^a - \text{Cov}^b\| \quad \leq \quad \|r_u^a\| + \|r_u^b\| \; .$$

对于具有多个评级的用户来说，这种限制可能很大，这导致我们将权重引入对协方差矩阵有贡献的项。 权重将被选择来仔细标准化每个用户的贡献，确保可能差异的标准至多是一个固定的常数。 因此，规范化后，我们可以简单地计算和报告协方差矩阵，并将固定的加性噪声量应用于每个条目。 正如我们所期望的那样，协方差矩阵中的值的大小随着数据点的数量呈线性增长，这种噪声的影响应随着训练数据量的增长而直观地减小。

4. 算法和分析

我们的算法由几个步骤组成，在将测量值送入当前顶级学习算法的适当参数化变体之前，先测量（带噪声）逐渐变得更具挑战性的数据。 在更具体地描述精确计算的序列之前，我们首先描述高层次的方法。

全局影响。 我们从各种全局影响的噪音测量和基线校正开始。 我们首先衡量并发布所有评级的总和和计数以得出全局平均值。 然后，我们为每部电影测量并发布该电影的评分数量和总和。 我们使用这两个数量来制作每部电影的平均值，通过在全局平均值中包含数量为β_m的评级来稳定。 最后，我们重新测量全局平均水平，如上所述，即将用于集中每个用户的评分。 这些步骤的算法和隐私含义在 4.2 节中详细描述。

接下来，我们在准备每个用户对协方差度量的评分时投入一些费用。 我们不希望释放每个用户的统计数据，例如每个用户的平均评分，因为如果足够的准确性对学习有用，这样做会破坏我们的隐私保证。 相反，我们会在测量之前对用户的评分进行几次转换，并认为转换是这样的，即其输出的隐私保证会传播到他们的输入中。 我们的具体操作包括将每位用户的评分集中在一起，再次包括全局平均评级数量，以及将结果值限制在更紧凑的时间间隔（增加隐私，以牺牲偏离值的误差为代价）。这些步骤的算法和隐私含义在 4.3 节中详细描述。

协方差矩阵。 接下来我们测量得到的用户评分向量的协方差矩阵。 为了实现隐私保护，我们将噪声合并到每个坐标中[7]。 作为有效整合隐私的微妙性质的一个例子，考虑从几何数据计算潜在因素。 许多几何学习方法中的一个重要步骤，我们可能想找到一个低维度

子空间最适合数据，当投影到它时，就均方误差而言。 计算这种空间有几种方法，但是另有三种等价的方法是计算用户×电影数据矩阵的 SVD，计算电影×电影协方差矩阵的 SVD，并计算用户×用户克拉矩阵的 SVD。

虽然这三种方法在非私人计算中是等价的，但在面对隐私整合任务时，它们却非常不同。考虑将噪声添加到测量中以提供隐私的简单技术：为了充分掩盖数据矩阵，我们必须在称为随机响应的过程中向矩阵中的每个条目添加噪声。尽管噪声的独立性导致了一定程度的取消，但系统中的错误仍然随着参与者的数量而增加。向协方差矩阵添加噪声随着参与的电影数量而变化，但不需要随着参与者数量的增加而增长，这为任意大量的人群提供了任意准确测量的可能性。使用格拉姆矩阵，每个用户对都有一个条目，这是一场灾难;必须为每个条目添加足够的噪声（参与者中的二次方），与任何两个用户可能具有的最大协方差成正比，电影数量是线性的。这很快就会变得无法控制的破坏性。

虽然这三种技术在没有隐私约束的情况下是相似的，但当我们需要为隐私引入噪声（协方差，数据矩阵，Gram 矩阵）时，它们之间有明确的排序。

4.1. 符号

与之前一样，r_{ui}代表用户 u 对电影 i 的评分，r_u代表用户 u 的整个评分向量，同样e_{ui}和e_u表示评估评分存在的二进制元素和向量（允许我们与报告的零区分值）。 我们使用c_u = || e_u || 1 表示用户 u 的评分数量。

在我们的展览中，我们将分别使用小写和大写来区分私人数据和发布数据。读者应该验证，无论何时分配一个大写字母变量，它只是一个只有大写字母变量或加入了噪音的小写字母变量的 函数。当我们将噪声添加到变量 x 时，我们只是写：

$$\mathrm{X} = x + \mathrm{Noise},$$

噪音分布尚未明确。 然后我们限制变量 x 在||·||下可以改变的量 1 和||·|| 2，并在数据集中增加一个额定值，允许增加拉普拉斯或高斯噪声，并分别通过定理 2 和 3 提供隐私保证。

4.2. 影片影响

我们从一些易于测量和发布的全局影响开始，不会产生大量的隐私成本。 我们首先衡量并公布每部电影的评分数量以及每部电影的总和或评分，并为隐私增加随机噪声：

$$\begin{aligned}\mathrm{GSum} &= \sum_{u,i} r_{ui} + \mathrm{Noise},\\ \mathrm{GCnt} &= \sum_{u,i} e_{ui} + \mathrm{Noise}.\end{aligned}$$

我们用这些来推导出一个全局平均值，G = GSum / GCnt。 接下来，我们使用 d 维矢量和来总结并计算每部电影的评分数量。

$$\begin{aligned}\mathrm{MSum} &= \sum_{u} r_u + \mathrm{Noise}^d,\\ \mathrm{MCnt} &= \sum_{u} e_u + \mathrm{Noise}^d.\end{aligned}$$

我们通过在每部电影中引入价值为 G 的β^m级别评级来产生稳定的每部电影平均评级：

$$\mathrm{MAvg}_i = \frac{\mathrm{MSum}_i + \beta_m \mathrm{G}}{\mathrm{MCnt}_i + \beta_m}.$$

随着现在公布的这些平均值，它们可以被随意地合并到随后的计算中，而不需要额外的隐私成本。 特别是，我们可以从每个评分中减去相应的平均值，以消除每部电影的全局影响。

4.3. 用户影响

发布每部电影的平均评分后，我们将在继续之前从每个评分中减去这些平均值。 然后，我们将每个用户的评分集中在一起，再次计算平均值，重新计算的全局平均值为β_p：

$$\bar{r}_u = \frac{\sum_i (r_{ui} - \mathrm{MAvg}_i) + \beta_p \mathrm{G}}{c_u + \beta_p}.$$

与电影不同，我们不报告平均数，我们只是从适当的评分中减去它们。 我们还会将得到的中心评级限制在区间[-B，B]中，以牺牲相对较少的剩余大项目来降低测量的灵敏度：

$$\widehat{r}_{ui} = \begin{cases} -B, & \text{if } r_{ui} - \overline{r}_u < -B, \\ r_{ui} - \overline{r}_u, & \text{if } -B \le r_{ui} - \overline{r}_u < B, \\ B, & \text{if } B \le r_{ui} - \overline{r}_u. \end{cases}$$

我们现在认为，存在或不存在单一评级对这种对中和夹紧过程有限制的影响。

定理 4.让r^a和r^b在一个等级上不同，用r^b表示。设 α是评分[2]中可能的最大差异。对于中心化和标准化的评级$\widehat{r^a}$和$\widehat{r^b}$，我们有：

$$\begin{aligned} \|\widehat{r}^a - \widehat{r}^b\|_1 &\le \alpha + B, \\ \|\widehat{r}^a - \widehat{r}^b\|_2^2 &\le \frac{\alpha^2}{4\beta_p} + B^2. \end{aligned}$$

证明。如果r^a和r^b是r^b中r^b_{ui}的单个新评级的两组评级，那么除了用户 u 的评级外，$\widehat{r^a}$和$\widehat{r^b}$在任何情况下都是相等的。对于r^a和r^b之间的共同评级，差异至多是相减的平均值中的差异：

$$|\overline{r}^b_u - \overline{r}^a_u| = \frac{|r_{ui} - \overline{r}^a_u|}{c^b_u + \beta_p} \le \frac{\alpha}{c^b_u + \beta_p}.$$

对于新的评级r_{ui}，其先前的零贡献被替换为新的居中和被限制的评级，设为最大值 B。因此：

$$\begin{aligned} \|\widehat{r}^a - \widehat{r}^b\|_1 &\le c^a_u \times \frac{\alpha}{c^b_u + \beta_p} + B, \\ \|\widehat{r}^a - \widehat{r}^b\|_2^2 &\le c^a_u \times \frac{\alpha^2}{(c^b_u + \beta_p)^2} + B^2. \end{aligned}$$

对于$\|\cdot\|_2^2$，*考虑到*c^b_u=c^a_u+ 1，第一项在c^a_u =βp+1 处最大，可以由α^2 /4βp 从上面限定。通过选择足够大的βp，我们可以驱动||·|| 2 差异任意接近于 B.对于||·|| 1 也是如此，我们只是用分母来取消它的c^a_u项。

4.4. 协方差矩阵

我们对私人数据的最终测量是中心和被限制的用户评分向量的协方差。但是，我们希望按照用户权重w_u等于"e_u"的倒数来非均匀地取平均值。规范的选择将决定我们派生稳定界限的规范。

$$\begin{aligned} \text{Cov} &= \sum_u w_u \widehat{r}_u \widehat{r}_u^T + \text{Noise}^{d\times d}, \\ \text{Wgt} &= \sum w_u e_u e_u^T + \text{Noise}^{d\times d}. \end{aligned}$$

请注意，如果单个评分r_{ui}在r^a和r^b之间有差异，则只有用户 u 贡献的术语会影响矩阵的差异。我们现在用定理 4 来限定这个差异的规范。

定理 5.让评级r^a和r^b有一个不同的评级。令w_u = 1 / || e_u || 1，我们有：

$$\|w^a_u \widehat{r}^a_u \widehat{r}^{aT}_u - w^b_u \widehat{r}^b_u \widehat{r}^{bT}_u\|_1 \le 2B\alpha + 3B^2 .$$

由于βp 至少为α^2/ 4B2，令w_u = 1 /||e_u ||2，我们有：

$$\|w^a_u \widehat{r}^a_u \widehat{r}^{aT}_u - w^b_u \widehat{r}^b_u \widehat{r}^{bT}_u\|_2 \le (1 + 2\sqrt{2})B^2 .$$

我们将差异值$w^a_u \widehat{r}^a_u \widehat{r}^{aT}_u - w^b_u \widehat{r}^b_u \widehat{r}^{bT}_u$ 改写为
$w^a_u \widehat{r}^a_u (\widehat{r}^a_u - \widehat{r}^b_u)^T + w^b_u (\widehat{r}^a_u - \widehat{r}^b_u) \widehat{r}^{bT}_u + (w^a_u - w^b_u) \widehat{r}^a_u \widehat{r}^{bT}_u.$

由于$\|\cdot\|_1 \text{ and } \|\cdot\|_2, \text{ as } \|e^b_u\| - \|e^a_u\| \le 1,$*我们有：*

$$w^a_u - w^b_u = \frac{1}{\|e^a_u\|} - \frac{1}{\|e^b_u\|} \le \frac{1}{\|e^a_u\|\|e^b_u\|} .$$

原始矩阵差异的范数受限于：

$$\left(\frac{\|\widehat{r}^a_i\|}{\|\widehat{e}^a_i\|} + \frac{\|\widehat{r}^b_i\|}{\|\widehat{e}^b_i\|}\right) \|\widehat{r}^a_i - \widehat{r}^b_i\| + \frac{\|\widehat{r}^a_i\|\|\widehat{r}^b_i\|}{\|\widehat{e}^a_i\|\|\widehat{e}^b_i\|}.$$

由于任何范数的||$\hat{r}_i$|| ≤|| $\hat{e}_i$|| ×B，归一化抵消了额定值的范数，通过定理 4 给出了补偿。

类似的结果适用于权重矩阵，但是可以在e_i矢量不进行居中时大体上优化。

定理 6.让评级r^a和r^b有一个不同的评级。令w_u= 1 / || e_u || 1，我们有：

$$\|w^a_u e^a_u e^{aT}_u - w^b_u e^b_u e^{bT}_u\|_1 \le 3 .$$

以w_u = 1 /||e_u||2，我们有：

$$\|w^a_u e^a_u e^{aT}_u - w^b_u e^b_u e^{bT}_u\|_2 \le \sqrt{2} .$$

证明。在两个权重矩阵之间，$(c^a_p)^2$项从w^a_p变为w^b_p，并且 2 c^b_p-1 项出现权重w^b_p。对于||·|| 1，这个界限是：

$$\|w^a_u e^a_u e^{aT}_u - w^b_u e^b_u e^{bT}_u\|_1 \le 3 - 2/c^b_p < 3.$$

对于||·|| 2 和c_p^a> 0 时，我们观察到在c_p^a：1/2$(c_p^a)^{3/2}$时，权重w_p^a-w_p^b的偏差至多是$x^{-1/2}$的导数，这意味着：

$$\begin{aligned}\|w_u^a e_u^a e_u^{aT} - w_u^b e_u^b e_u^{bT}\|_2^2 &\leq \frac{(c_p^a)^2}{4(c_p^a)^3} + \frac{(2c_p^b - 1)}{(c_p^b)^2} \\ &< 2.\end{aligned}$$

通过直接计算来处理c_p^a= 0 的情况。

4.5. 任一用户隐私

到目前为止，我们所做的数学描述了掩盖是否存在单个评分所需的噪音量。 更强大的隐私保证将掩盖整个用户的存在或缺失，即使对于超级电影评分者也提供统一的隐私保证。 为了更新数学以提供每个用户的隐私，我们只需要按照评分数量应用更积极的降低权重，将每个贡献缩小为|| e_u ||。 对于总和和计数的贡献，新用户准确地贡献他们的加权评分和计数向量：

$$\|\frac{r_u}{\|e_u\|}\| \leq \alpha \quad \text{and} \quad \|\frac{e_u}{\|e_u\|}\| \leq 1.$$

同样，对协方差和权矩阵的贡献正好是加权向量的新外积，其规范是向量范数的平方：

$$\|\frac{\widehat{r}_u}{\|e_u\|}\|^2 \leq B^2 \quad \text{and} \quad \|\frac{e_u}{\|e_u\|}\|^2 \leq 1.$$

这种标准化比每个级别的隐私更具侵略性，并导致不太准确的预测和推荐。 但是，即使随着用户数量的增加，噪声量仍然保持固定。

4.6. 净化协方差矩阵

我们计算出的协方差矩阵有点嘈杂，虽然我们可以将它交给许多推荐算法之一，但我们首先将它清理一下。

作为第一步，我们将“收缩到平均值”方法[4]应用于对角线条目和对角线条目的单独常量，将矩阵设置为：

$$\overline{\text{Cov}}_{ij} = \frac{\text{Cov}_{ij} + \beta \cdot \text{avg Cov}}{\text{Wgt}_{ij} + \beta \cdot \text{avg Wgt}}.$$

有大量的理论和经验证据表明，低秩矩阵逼近（与矩阵分解方法相同的形式）对于从矩阵去除噪声同时保持显着的线性结构是非常有效的。 通过计算我们的协方差矩阵的 rank-k 近似值，我们可以消除我们已经引入的大量“平方误差”，而不会从底层信号中去除几乎相同的误差。

在应用 rank-k 近似之前，我们希望尽可能统一噪声的方差。 贡献项相对较少的协方差条目在其增加的噪音方面具有较高的差异（因为它被分开

由一个较小的Wgt_{ij}）。 也有人观察到，当条目的方差是等价的时候，低秩近似的去噪是最有效的（误差界限为第一个近似值，按照每个条目的方差最大的比例，并且不会导致放大 较少的条目，同时增加每个“信号”的贡献量）。

为了纠正这个问题，我们借用了[12]中的技术，并在应用 rank-k 近似值之前将每个入口Avg_{ij}向上缩放$(MCnt_i MCnt_j)^{1/2}$。 然后，我们用相同的因子缩小每个条目，并将结果返回给我们的推荐算法。 这一步的另一个好处是它可以生成高度压缩的协方差矩阵，现在可以将其全部发送到客户端计算机。

5. 评估

我们在 Netflix 奖数据集上评估我们的方法，该数据集由 480,000 人贡献的约 17700 部电影的大约 100M 评级组成。 通过调整我们使用的噪声分布的参数，我们的计算将提供不同的差分隐私保证，并且其输出将具有可测量的准确性属性。 精度是通过限定集（3M 评分）上的均方根误差（RMSE）来测量的，并且可以在具有相似特征（1.5M 评分）的指针组上进行自我测试。

5.1. 隐私与准确性交易

尽管使用各种(ε, δ)对对参数化差分隐私是很自然的，但我们简化为单个参数。 对于每个测量f_i，我们将参数化我们用作噪声的幅度：

$$\sigma_i = \max_{A \approx B} \|f_i(A) - f_i(B)\|/\theta_i \ ,$$

其中θ_i需要总和为预先指定的值θ。 事实上，我们将把每个θ_i都作为θ的一个固定分数，我们将采用这个值作为我们的单一参数。 通过定理 2，使用拉普拉斯噪声，测量结果为我提供了ε_i-差分隐私：

$$\epsilon_i = \theta_i.$$

根据定理 3，使用高斯噪声，测量值 i 提供了(ε_i, δ_i) – 差分隐私。

$$\epsilon_i \quad = \quad \theta_i \sqrt{2\ln(2/\delta_i)}.$$

由于定理 1 告诉我们ε和δ值相加，我们的最终保证有形式（分别为||·|| 1 和||·|| 2）

$$\epsilon \;=\; \sum_i \theta_i \quad \text{and} \quad \epsilon \;=\; \sum_i \theta_i \sqrt{2\ln(2/\delta_i)}$$

通过采用δ_i的常见值，我们可以看到θ= $\sum \theta_i$线性缩放了ε。 通过改变θ，从而改变θ_i，我们可以为我们的测量增加更多或更少的噪声，并分别提供更多或更少的隐私。 从任何θ，我们都可以重构(ε, δ)对的范围。

第 4 节描述了全局效应和协方差矩阵的隐私保护计算。 我们的算法有三个重要的测量数据：全局平均值，每部电影平均值和协方差矩阵。 对于任何θ，我们将根据其设置相应的θ_i

$$\theta_1 \;=\; 0.02 \times \theta, \quad \theta_2 \;=\; 0.19 \times \theta, \quad \theta_3 \;=\; 0.79 \times \theta\,.$$

我们选择θ_1这么小，因为每个评分都有助于其计算，即使有大量的加性噪声，所得到的平均值也是非常准确的。

考虑到协方差矩阵和全局效应的因素，我们应用 Bell 和 Koren 的 k-Nearest Neighbor（kNN）方法和基于 SVD 的标准预测机制（均采用岭回归）。 我们使用权重矩阵W_{gt}作为 kNN 的相似性度量。 这两种方法都可以在清洁步骤之前进行（第 4.6 节），这可能会根据隐私参数的值而改善或降低性能。

所有全局参数均针对θ= 0.15 的值进行了优化，其中 SVD 和 kNN 均在清洁步骤之前与 Netflix 的 Cinematch 基准相匹配。 所有算法的维数固定为 k = 20，收缩参数β_m = 15 和β_p = 20，钳位参数 B = 1.0。 清洁步骤以及基于 kNN 和 SVD 的推荐机制的参数分别针对每个数据集和隐私参数进行训练，因为可以使用相对较少的取决于私人数据的新测量来完成。 我们使用梯度下降法，它使用不同的参数反复评估每个机构。 这种配置不需要用来重新测量协方差矩阵，所以我们不会在那里产生额外的隐私成本。 但是，我们必须评估 RMSE，即使隐私参数设置非常低，也可以以极好的精度进行测量。

我们的主要研究结果如图 1 所示。由于与σ成反比的θ值和隐私质量的数量增加，推荐算法的准确性也随之增加。 k-NN 和 SVD（均带清洗）均在θ≈0.15 时与 Cinematch 阈值相交。 如果没有清洗，k-NN 和 SVD 都会通过基线，但噪音较小，因此隐私性较差。 尽管协方差矩阵的后处理"清理"在很大程度上有助于高噪声（小θ）状态，但当使用较少的噪声时会损害分析。 这可能是优化θ= 0.15 的清洁参数的结果，并且可能的是，更精细的后处理可以适应两种制度。

本文的完整版本中出现了拉普拉斯噪声和用户级隐私的相应图表。

5.2. 隐私与时间的准确性

我们的算法（如大多数差分隐私计算）为任何测量引入了固定的误差量，随着数据集大小的增加，这种误差越来越受实际数据记录的支配。 随着越来越多的用户和评级，我们期望我们为任何固定的θ值引入的加性误差最终消失。

为了探索由于推荐机制的隐私保护特性造成的损失如何随着可用数据量的减少（对于固定值θ= 0.15），我们模拟了 2000 年至 2006 年间不同时间的数据收集过程（包括评分少于 20 的用户的特有属性）。 与 Netflix Prize 数据集一致，探针组是每个用户选择的最近 9 个评级，每个用户的概率为 1/3。 图 2 绘出了隐私保护 k-NN（缩放后）和没有隐私保护的相同算法之间 RMSE（以百分点表示）的差异。

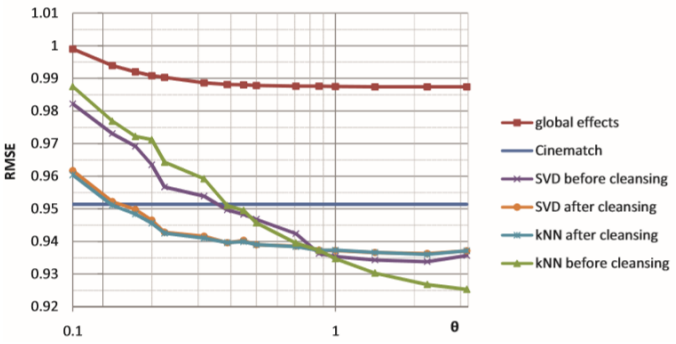

图 1：四种算法的 RMSE 作为 Net fl ix Prize 集上θα1 / σ的函数。

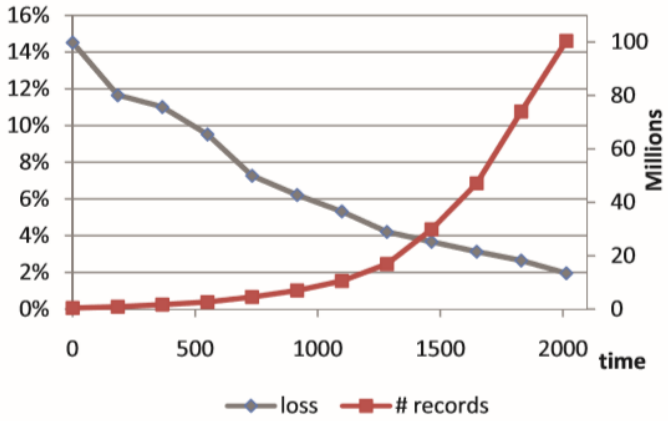

图 2：左尺度精度损失，右尺度 - 记录数量。 x 轴是 2000 年 7 月 1 日以来经过的天数。

6. 结论和未来的工作

我们的结论是，具有差分隐私保证的推荐系统是可行的，而不会对建议的准确性产生重大影响。 随着更多数据可用，准确性（对于隐私参数的固定值）的损失减少。 在我们的实验中，我们固定了几个可以自由变化的参数，期望更深入的实验能够显着提高预测精度是自然的。 所选择的维度，平滑权重和测量之间的“准确度”θ_i的分布可以被调整并且可能被改进。 未来工作的方向包括直接隐私保护计算潜在因素的有效方法，并纳入协作过滤高级方法的差异性隐私框架，这种方法不会立即将因式分解分为两个阶段，如[19]的综合模型。

7. 引用

[1] C. C. Aggarwal. On k-anonymity and the curse of dimensionality. In K. B¨ohm, C. S. Jensen, L. M. Haas, M. L. Kersten, P.-˚A. Larson, and B. C. Ooi, editors, VLDB, pages 901–909. ACM, 2005.

[2] E. A¨ımeur, G. Brassard, J. M. Fernandez, and F. S. M. Onana. Alambic: a privacy-preserving recommender system for electronic commerce. Int. J. Information Security, 7(5):307–334, 2008.

[3] E. A¨ımeur, G. Brassard, J. M. Fernandez, F. S. M. Onana, and Z. Rakowski. Experimental demonstration of a hybrid privacy-preserving recommender system. In ARES, pages 161–170. IEEE Computer Society, 2008.

[4] R. M. Bell and Y. Koren. Scalable collaborative filtering with jointly derived neighborhood interpolation weights. In ICDM, pages 43–52. IEEE Computer Society, 2007.

[5] R. M. Bell, Y. Koren, and C. Volinsky. The BellKor solution to the Netflix Prize. Available from http://www.netflixprize.com, 2007.

[6] R. M. Bell, Y. Koren, and C. Volinsky. The BellKor 2008 solution to the Netflix Prize. Available from http://www.netflixprize.com, 2008.

[7] A. Blum, C. Dwork, F. McSherry, and K. Nissim. Practical privacy: the SuLQ framework. In C. Li, editor, PODS, pages 128–138. ACM, 2005.

[8] J. Brickell and V. Shmatikov. The cost of privacy: destruction of data-mining utility in anonymized data publishing. In Li et al. [20], pages 70–78.

[9] J. Calandrino, A. Narayanan, E. Felten, and V. Shmatikov. Don’t review that book: Privacy risks of collaborative filtering. Manuscript, 2009.

[10] J. F. Canny. Collaborative filtering with privacy. In IEEE Symposium on Security and Privacy, pages 45–57, 2002.

[11] J. F. Canny. Collaborative filtering with privacy via factor analysis. In SIGIR, pages 238–245. ACM, 2002.

[12] A. Dasgupta, J. E. Hopcroft, and F. McSherry. Spectral analysis of random graphs with skewed degree distributions. In FOCS, pages 602–610. IEEE Computer Society, 2004.

[13] C. Dwork. Differential privacy. Invited talk. In Automata, Languages and Programming—ICALP (2), volume 4052 of Lecture Notes in Computer Science, pages 1–12. Springer, 2006.

[14] C. Dwork. An ad omnia approach to defining and achieving private data analysis. In F. Bonchi, E. Ferrari, B. Malin, and Y. Saygin, editors, PinKDD, volume 4890 of Lecture Notes in Computer Science, pages 1–13. Springer, 2007.

[15] C. Dwork. Differential privacy: A survey of results. In M. Agrawal, D.-Z. Du, Z. Duan, and A. Li, editors, Theory and Applications of Models of Computation, 5th International Conference, TAMC 2008, Xi’an, China, April 25-29, 2008. Proceedings, volume 4978 of Lecture Notes in Computer Science, pages 1–19. Springer, 2008.

[16] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. Our data, ourselves: Privacy via distributed noise generation. In S. Vaudenay, editor, Advances in Cryptology—EUROCRYPT 2006, volume

4004 of Lecture Notes in Computer Science, pages 486–503. Springer, May 2006.

[17] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In S. Halevi and T. Rabin, editors, Theory of Cryptography Conference—TCC 2006, volume 3876 of Lecture Notes in Computer Science, pages 265–284. Springer, 2006.

[18] S. R. Ganta, S. P. Kasiviswanathan, and A. Smith. Composition attacks and auxiliary information in data privacy. In Li et al. [20], pages 265–273.

[19] Y. Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In Li et al. [20], pages 426–434.

[20] Y. Li, B. Liu, and S. Sarawagi, editors. Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, Nevada, USA, August 24-27, 2008. ACM, 2008.

[21] A. Machanavajjhala, D. Kifer, J. M. Abowd, J. Gehrke, and L. Vilhuber. Privacy: Theory meets practice on the map. In ICDE, pages 277–286. IEEE, 2008.

[22] A. Narayanan and V. Shmatikov. Robust de-anonymization of large sparse datasets. In IEEE Symposium on Security and Privacy, pages 111–125. IEEE Computer Society, 2008.

[23] L. Sweeney. k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(5):557–570, 2002