现代消费者淹没了选择。电子零售商和内容提供商提供了大量的产品，具有前所未有的机会满足各种特殊需求和品味。消费者与最适合的产品匹配是提升用户满意度和忠诚度的关键。因此，更多的零售商已经对推荐系统感兴趣，推荐系统分析用户对产品感兴趣的模式，以提供适合用户口味的个性化推荐。由于良好的个性化建议可以为用户体验添加另一个维度，像Amazon.com和Netflix这样的电子商务领导者将推荐系统作为其网站的重要组成部分。

这样的系统对诸如电影，音乐和电视节目的娱乐产品特别有用。许多客户将会看到同一部电影，每个客户都可能会看到许多不同的电影。客户已经证明愿意表明他们对特定电影的满意程度，所以大量的数据可用于哪些电影吸引哪些客户。公司可以分析这些数据，向特定客户推荐电影。

# 推荐系统

广义而言，推荐系统是基于两种策略之一。内容过滤方法为每个用户或产品创建一个配置文件，以表征其性质。例如，电影简介可以包括关于其类型，参与演员，票房流行度等的属性。用户个人资料可能包含人口统计信息或在适当问卷上提供的答案。配置文件允许程序将用户与匹配的产品相关联。当然，基于内容的策略需要收集可能不可用或易于收集的外部信息。已知的成功实现内容过滤是音乐基因组计划，用于互联网无线电业务Pandora.com。

内容过滤的替代方案仅依赖于过去的用户行为 - 例如，以前的交易或产品评级，而不需要创建显式配置文件。这种方法被称为协作过滤，这是由Tapestry的开发人员创建的术语，第一个推荐系统。协同过滤分**析用户之间的关系和产品之间的相互依赖关系**，以确定新的用户项目关联。合作筛选的主要吸引力在于它是免费域名的，但它可以解决通常使用内容过滤难以实现和难以概述的数据方面。虽然通常比基于内容的技术更准确，但协作过滤由于无法解决系统的新产品和用户而导致所谓的冷启动问题。在这方面，内容过滤是优越的。

协同过滤的两个主要领域是**邻域方法和潜在因素模型**。邻域方法的核心是计算项目之间或者在用户之间的关系。面向对象的方法基于相同用户对“相邻”项目的评级来评估用户对项目的偏好。产品的邻居是在同一用户评级时倾向于获得类似评分的其他产品。例如，考虑电影Saving Private Ryan。其“邻居”可能包括战争电影，斯皮尔伯格电影和汤姆·汉克斯电影等。要预测特定用户对“Saving Private Ryan”的评级，我们将寻找该用户实际评分的电影最近的邻居。如图1所示，面向用户的方法确定了可以补充对方评级的志同道合的用户。

潜在因素模型是一种替代方法，尝试通过表征项目和用户来解释评级，例如，从评级模式推断出20到100个因素。在某种意义上，这些因素包括与上述人造歌曲基因的计算机替代。对于电影，发现的因素可能衡量明显的维度，如喜剧与戏剧，行动量或儿童的方向; 不太明确的维度，如角色发展或奇怪的深度; 或完全不可解释的尺寸。对于用户而言，每个因素测量用户喜欢在相应的电影因素中得分高的电影。

图2说明了两个维度中简化示例的想法。考虑两个假设的层面定性为女性与男性为导向，认真与逃避现实。该图显示了几个知名电影和一些虚构的用户可能落在这两个方面。对于该模型，用户对电影的预测评级相对于电影的平均评级，将等于电影和用户在图表上的位置的点积。例如，我们希望Gus爱“阿呆与阿瓜”，讨厌“紫色姐妹花”，并评价“勇者之心”的平均水平。请注意，一些电影 - 例如，“十一罗汉” - 和用户 - 例如，戴维 - 将在这两个维度上被表征为相当中立。

# 矩阵分解方法

潜在因素模型的一些最成功的实现是基于矩阵分解。在其基本形式中，矩阵因式分解是通过从项目评级模式推断的因子的向量来表征项目和用户。项目和用户因素之间的高对应性导致了一个建议。近年来，通过结合良好的可扩展性和预测精度，这些方法已经变得流行。此外，它们为建模各种现实生活中的场景提供了很大的灵活性。

推荐系统依赖于不同类型的输入数据，这些数据通常放置在矩阵中，一维代表用户，另一维度表示感兴趣的项目。最方便的数据是高质量的明确反馈，其中包括用户对产品兴趣的明确输入。例如，Netflix收集电影的星级，TiVo用户通过按竖起大拇指按钮和拇指按钮来表示他们对电视节目的偏好。我们将明确的用户反馈称为评级。通常，显式反馈包括稀疏矩阵，因为任何单个用户可能仅对可能项目的一小部分进行评估。

矩阵分解的一个优点是它允许并入额外的信息。当显式反馈不可用时，推荐系统可以使用隐含反馈来推断用户偏好，通过观察用户行为（包括购买历史，浏览历史，搜索模式，甚至鼠标移动）来间接反映意见。隐性反馈通常表示事件的存在或不存在，因此通常由密集填充的矩阵表示。

# 一个基本的矩阵分解模型