

## 081 | 基于深度学习的推荐模型之一：受限波兹曼机

2018-04-09 洪亮劼

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 07:21 大小 3.37M



这几周我们进入专栏里一个比较大的模块，那就是推荐系统。

上周，我们谈了现代推荐系统的架构体系，帮助你在宏观上对推荐系统的构建有一个更加完整的认识。这周，我们主要来看在推荐系统研究领域里一个比较前沿的话题，那就是**如何利用深度学习来提升推荐系统的精度**。

今天，我们首先来看一篇经典的文章《受限波兹曼机进行协同过滤》（[Restricted Boltzmann machines for collaborative filtering](#)）[1]，这篇文章尝试使用**受限波兹曼机**（Restricted Boltzmann Machines），简称 **RBM**，来对推荐系统进行建模。这应该是最早把深度学习应用到推荐建模的典范。

### 受限波兹曼机（RBM）

从严格意义上讲，RBM 自身并不是深度模型，在早期对于 RBM 的使用上，也并没有将其“累加”到很多层从而形成“深度 RBM”（Deep RBM）。但是从建模思路上来说，由一层 RBM 到多层 RBM 的扩展其实是非常直接的，因此了解 RBM 的基本思路，对后面理解推荐系统中如何利用深度学习模型进行建模是很有帮助的。

今天我们要介绍的文章发表于 2007 年的 ICML（International Conference on Machine Learning，国际机器学习大会）上。作者群是鲁斯兰·萨拉胡特蒂诺夫（Ruslan Salakhutdinov）、安德烈·尼哈（Andriy Mnih）以及杰弗里·辛顿（Geoffrey Hinton）。

辛顿在近日常常被称作深度学习之父。他不仅算是波兹曼机（Boltzmann Machines）的重要发明人和推动者之一，也和其学生一起把 RBM 应用到各类数据上，比如 RBM 在推荐领域的尝试。

2007 年的时候，Netflix 大赛如火如荼，很多学者都希望把各种模型和思路应用到这个比赛中。而在这个大赛中，有三个重要的思想脱颖而出，影响了后来推荐系统的研究发展。这三个思想分别是：

1. 基于矩阵分解的模型；
2. 基于 RBM 的模型；
3. 利用集成学习（Ensemble Learning）把大量不同的模型整合起来。

由此可见，RBM 对于推荐系统的尝试在当时是非常有新意的。

第一作者鲁斯兰当时在多伦多大学攻读计算机博士，跟随辛顿研究深度学习模型。另外一篇他当时做的工作，是把贝叶斯矩阵分解利用到 Netflix 大赛中，和我们今天讨论的这篇论文一起，都是他在博士阶段对于推荐系统这个领域所做的重要贡献。目前鲁斯兰在卡内基梅隆大学任教，并兼任苹果公司的人工智能总监一职。

那什么是 RBM 呢？简单说来，RBM 就是由一层**隐单元**（Hidden Unit）和一层**显单元**（Visible Unit）组成的神经网络结构。通常情况下，显单元和隐单元这两层之间是完全连通的。也就是说，对于每一个显单元来说，它都和所有的隐单元联系到一起。

**每个隐单元和显单元自身都有一个权重（Weight），并且在每对隐单元和显单元之间的连接上还有一个权重。**所有这些权重都是需要通过训练学习的未知参数。举例来说，如果我们

有 3 个显单元（用于描述 3 个数据点），5 个隐单元。那么我们就有 3 个权重对应 3 个显单元，有 5 个权重对应 5 个隐单元，然后有 15 ( $3 \times 5$ ) 个连接权重，这样算下来一共是 23 个权重。RBM 可以针对高斯信号，也就是实数信息，以及伯努利或者二项分布 (Binomial Distribution) 信号，也就是离散信息，进行建模。

## 受限波兹曼机在推荐上的应用

当我们对 RBM 有了一个基本的了解之后，我们来看 RBM 是怎么应用到推荐系统上的。为了讲解方便，我们这里使用 Netflix 的例子，也就是对“用户为电影打分”进行建模。

首先，对于每一个用户，我们使用一个单独的 RBM 来对这个用户进行建模，只不过，每一个 RBM 都有一样的隐单元数量。在建模的时候，每一个 RBM 仅仅对当前这个用户曾经打过分的数据进行建模。也就是说，每一个 RBM 需要建模的数据是不一样的，有的用户对三个电影打过分，有的对十个电影打过分。你可以设想一下，在这样的情况下，整个模型是什么意思。

每个用户有一个独立的 RBM，这个 RBM 负责对这个用户的电影集合进行建模。很显然，这些 RBM 互相没有关联。那怎么把每个用户的 RBM 给联系起来呢？作者们做了这样的假设，也就是每个 RBM 的隐单元是不一样的，这其实可以代表学习到的“用户的偏好”。但是，如果两个用户都对同一个电影打过分，那么，针对这个电影，两个不同的 RBM 会共享一样的权重。这就是联系两个 RBM 的核心思想，也就是说，**利用 RBM 对用户电影推荐的核心是“共享相同电影的权重”**。

具体说来，每一个显单元都是用户对于某个电影的评价。这里每个显单元都是一个 K 维的数组，其中只有一个元素是 1，其他都是 0。这种**二元的数组表达**，帮助模型来对“用户对于有 K 种可能的输出”进行建模。隐单元在这篇论文中，也是二元的，只不过，我们事先并不知道隐单元的取值。和刚才介绍的一样，在这样的模型里，需要学习的参数包括显单元的权重、隐单元的权重以及他们之间关系的权重，同一个电影的权重是共享的。

每一个用户都有一个单独的 RBM，并且我们在 RBM 里只对已经评过分的电影进行了建模，因此这个模型并不能直接对未知的电影评分进行预测。需要预测的时候，我们其实是先拿到这个电影的权重，然后看，我们把那个 K 维的评分数组设置成哪种情况的时候，产生的概率是最大的。也就是说，我们尝试把对于未知的电影评分设置成不同情况，取出现可能性最大的那种评分作为预测结果。很明显，这样做的计算效率并不高。文章中也介绍了如何加速，我们这里就不复述了。

**RBM 对于推荐系统的建模看上去很简单，但是难点却是如何学习这些未知的权重。**总体说来，RBM 无法直接使用类似“最大似然法”或者“递归下降”的方法来对参数进行学习。这其实也是阻碍这类方法广泛使用的一个重要障碍。在最初的论文里，作者们提出了一种 CD (Contrastive Divergence, 对比散度) 方法，这种方法其实是一种简化的 MCMC (Markov chain Monte Carlo, 马尔科夫蒙特卡洛) 方法，用于对 RBM 进行采样。CD 方法具体是如何实现的，我们这里就不展开了。

在 Netflix 的数据集上，RBM 展现了惊人的效果，不仅能够很轻松地击败 Netflix 自身的算法基线，还比当时提出的单纯的矩阵分解方法要更加优秀。基于此，后来就有了很多 RBM 的扩展工作。

## 小结

这周我来为你讲解利用推荐系统的一个重要的问题，就是如何利用深度学习模型来对推荐系统进行建模。今天我们聊了一个最基本的深度学习模型，受限波兹曼机，讨论了如何将其应用在推荐系统中。

一起来回顾下要点：第一，我们介绍了什么是受限波兹曼机；第二，我们讨论了如何把受限波兹曼机应用到推荐场景中。

最后，给你留一个思考题，如果希望在受限波兹曼机里增加其他信息，比如各种用户信息或者电影信息，我们该如何做呢？

欢迎你给我留言，和我一起讨论。

## 参考文献

1. Ruslan Salakhutdinov, Andriy Mnih, and Geoffrey Hinton. Restricted Boltzmann machines for collaborative filtering. Proceedings of the 24th international conference on Machine learning (ICML '07), Zoubin Ghahramani (Ed.). ACM, New York, NY, USA, 791-798, 2007.

---

# AI 技术内参

你的360度人工智能信息助理

洪亮劼

Etsy 数据科学主管  
前雅虎研究院资深科学家



新版升级：点击「👤 请朋友读」，10位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 080 | 现代推荐架构剖析之三：复杂现代推荐架构漫谈

下一篇 082 | 基于深度学习的推荐模型之二：基于RNN的推荐系统

## 精选留言 (1)

写留言



林彦

2018-04-09



可以通过条件受限玻尔兹曼机引入一个表示用户特征，并把它与隐单元建立全连接。还有通过朴素贝叶斯引入用户和电影特征的。