

基于深度学习推荐系统的研究

刘春阳, 申莹, 赵永翼

(沈阳师范大学, 沈阳 110034)

摘要: 针对深度学习的研究, 仅运用单一的传统算法在数据稀疏的情况下, 不能得到隐式的反馈。所以提出了推荐系统, 在机器学习上有一定的影响, 并在社会上的应用也非常广泛, 在最近几年在图像处理、会话识别、对象侦探以及药物发现和基因组学等方面取得了一定的发展。将深度学习应用到推荐系统中, 探究怎样将大量的数据转变成多源异构数据, 它能够让我们对用户在使用过程中得到整合数据, 得出用户的偏好。优化推荐系统, 了解用户喜好, 是我们主要研究目的。在推荐系统系统中, 我们主要针对协同过滤进行研究。

关键词: 推荐系统; 深度学习; 协同过滤; 多源异构数据

doi: 10.3969/J.ISSN.1672-7274.2020.01.072

中图分类号: TP3

文献标识码: A

文章编号: 1672-7274 (2020) 01-0103-02

0 引言

在最近几年, 随着互联网在各个领域的发展, 将人类带入了快速网络经济时代, 互联网+在各个领域的应用表现出很大优势, 大数据在应用中有着很大的优势, 会带给人类在网络社会飞速的发展, 但也带来了很多问题, 比如“信息过载”的问题, 所以用户不得不从浩如烟海的信息海洋获取真正需求的信息。推荐系统能够成为缓解“信息过载”最有效的方法之一, 已经应用到电子商务, 电影, 视频等很多领域。并取得了很大的进步。在个领域也有了一定的成果。我们利用协同过滤的方式的矩阵分解对用户构建喜好矩阵。得到隐性向量, 组成推荐模型, 其研究具有很大的学术价值以及商业应用价值, 目前最经典的是协同过滤算法, 同时也是应用最广泛的算法。

1 深度学习技术—受限玻尔兹曼机

玻尔兹曼机是一种神经网络, 由 Hinton 和 Ackley 两位学者作为研究学者在不同的领域不同的目的同时提出的 BM 学习机, 通过从训练样本得到的隐层神经元状态上执行马尔可夫链抽样过程, 隐藏的神经元和可见的神经元是组成 BM 的主要神经元。BM 有很强的学习能力主要是能够学习数据规则, 且较为复杂的。但是, 受限玻尔兹曼机的训练时间很长。因为隐藏层神经元独立于给定的训练样本并且神经元之间不连接。这也让从属数据的期望值变得容易。其状态为 1-2, 状态 1 表示抑制下的神经元, 状态 2 表示激活下的神经元。因此提出了一种受限玻尔兹曼机的结构。如图 1 所示, 包括可见层 v 和隐藏层 h , 两层节点上下都是连接的, 而左右相互之间没有连接。

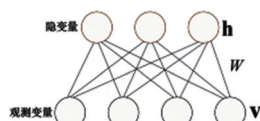


图1 受限玻尔兹曼机示意图

2 深度学习在协同过滤中的应用

我们通过学习深度学习学习到使用者表示的隐性和向量的低维矢量来实现建议的协同过滤算法经常面临可伸缩性不足的问题。但他们中的大多数现在推荐协同过滤应用程序。

我们利用使用者反馈给我们的隐式数据经过深度学习让模型输出等于输入。但要保证得到的隐式数据维度要一致, 提取数据进行神经协同过滤, 将隐性数据串联到一起组成使用者与项目关系。提高准确性, 利用辅助信息, 可以将推荐信息推荐给使用者, 使用户能够在海量的数据中找到准确的信息。

基于受限玻尔兹曼机的协同过滤方法

受限玻尔兹曼机是玻尔兹曼机的一种特殊拓扑结构, 基于 RBM 的协同过滤。推荐系统传统的定义可以理解为采集用户信息, 结合推荐模型帮助用户选择商品或者提供建议的过程。这就需要数据得收集, 预处理, 推荐算法和产生推荐 3 部分组成。例如一个典型的电影推荐系统一般包括含有 N 个用户的用户集合 $U = \{u_1, u_2, u_3, \dots, u_N\}$ 和含有 M 个项目的项目集合 $I = \{i_1, i_2, i_3, \dots, i_M\}$, 使用者在对 I 评价时只有一小部分, 这一部分的会表示成 RNM , 得到的数据稀疏, 所以推荐的信息不够准确, 因此, 我们采用一个 0-1 向量 r 表示电影是否被评过, 通过加入这种辅助数据得到了一种条件 RBM。模型的结构如图 2 所示。

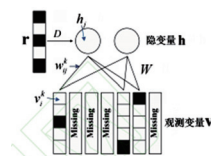


图2 基于RBM的协同过滤方法的模型结构

依据 RBM 模型可见单元和隐层单元之间的独立性质, 当给定可见状态时, 可见层单元 A 和隐层单元 C 的条件概率可以表示为:

$$p(a_i^k = 1 | h) = \frac{\exp(b_i^k + \sum_{j=1}^F c_j w_{ij}^k)}{\sum_{i=1}^K \exp(b_i^k + \sum_{j=1}^F c_j w_{ij}^k)}$$

$$p(c_j = 1 | a, r) = \sigma(b_j + \sum_{i=1}^m \sum_{k=1}^K c_i^k w_{ij}^k + \sum_{i=1}^m r_i D_y)$$

式中, K 和 F 表示用户和隐变量的数量。

3 结束语

本文对深度学习应用于推荐系统分析与研究, 可以知道深度学习不仅可以应用到推荐系统中, 还可以运用到各个领域, 本文将深度学习技术融入推荐系统中, 讨论了推荐算法的分类及算法的思想, 并着重分析了协同过滤算法, 阐释推荐算法所存在的问题, 试验方法, 和评测指标, 通过对协同过滤算法的研究, 可以反映出不同的用户不同的偏好度, 同时也可以提高推荐的精确度。现在, 在很多研究中使用深度学习来解决问题, 尽管目前信任是个问题, 但它终将被解决。我相信在未来深度学习会成为一项很重要的技术手段。

参考文献 (下转第 120 页)

作者简介: 刘春阳, 女, 1993 年生, 辽宁沈阳人, 研究生在读, 主要研究方向为数据挖掘。

申莹, 女, 1991 年生, 辽宁营口人, 研究生在读, 主要研究方向为数据挖掘。

赵永翼, 男, 汉族, 1972 年生, 辽宁沈阳人, 博士研究生, 教授, 主要研究方向为数据仓库与数据挖掘。

根据以上仿真,即可得到分析区域内的所有建筑物内部的覆盖情况。由于采用的是室外宏站覆盖室内的仿真方法,因此需要剔除已完成室内分布系统的建筑物。

将已完成室内覆盖的建筑物信息标记在 Googleearth 或者其他地图软件中,并对仿真数据进行对比,室内已覆盖建筑均采用红色标示,具体如图2所示。



图2 已完成室内覆盖建筑物标记

根据对比结果,建筑物室内弱覆盖且无室内分布系统的情况均可识别到。将对比得到的室内弱覆盖建筑进行地理化标示,整理并输出结果即可得到现网中室内弱覆盖的汇总情况。具体分析方法可以参照图3所示。

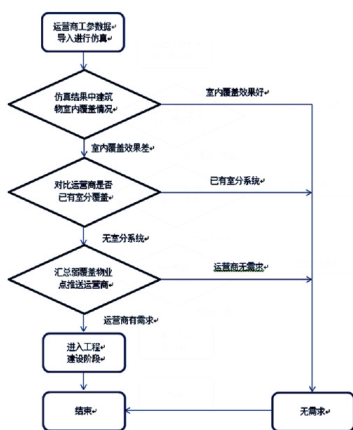


图3 室内弱覆盖分析方法流程

2.3 室内弱覆盖仿真分析方法验证

为了更好的验证此种分析方法的可行性,以及仿真效果和现网覆盖状况的吻合度,对河南省郑州市进行部分区域室内覆盖仿真,并对仿真中存在弱覆盖且当前无室内覆盖的建筑物进行室内路测验证。

选取郑州市某商务楼,仿真工程显示该建筑物内存在弱覆盖严重,同时结合现网室内分布系统建设情况,该建筑并未进行室内覆盖建设。根据仿真工程对其3、4、6层进行仿真数据抽取,并对这三层进行室内路测,室内路测与仿真对比如图4所示。路测的 RSRP 值与仿真数据对比如图5所示,两者吻合度较高。本

次仿真中采用了50%的网络负荷指标,而实际中现网用户数量及业务量均不高,两者差异会相对较大,因此,本次未对比 RS-SINR 数据指标。

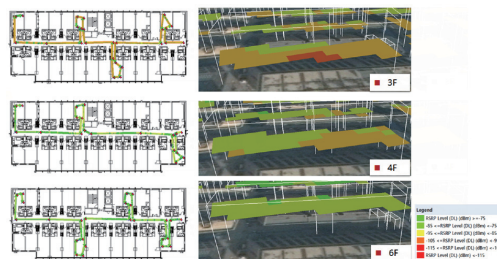


图4 某商务楼3、4、6层路测及仿真数据

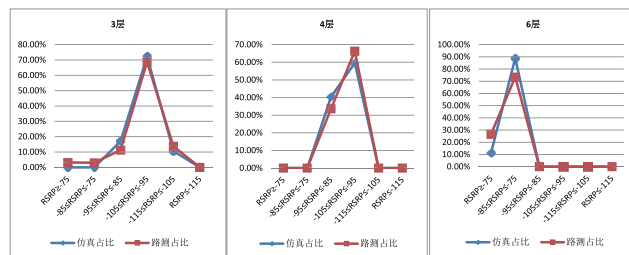


图5 某商务楼3、4、6层仿真及路测数据RSRP对比

对比仿真与路测数据,就覆盖指标而言,基于 CrossWave 射线传播模型所进行的室内弱覆盖仿真与现网吻合度较高,具备快速分析现网室内弱覆盖的基本要求。

3 结束语

本文根据 CrossWave 射线传播模型实现了室外覆盖室内的仿真,并进行现网室内路测验证,充分证明了基于 CrossWave 射线传播模型可以快速实现室内弱覆盖的批量定位,不仅降低了网络优化测试人员现场勘查测试时间和频次,同时也提高了获取室内弱覆盖的数量,因此基于 CrossWave 射线传播模型的室内弱覆盖分析方法对工程开始前期的拉网测试具有重要的指导意义,提高了网络优化人员分析工程覆盖目标的效率。

参考文献

- [1] 雷鸣,黄小光,汪伟,等. CrossWave 射线模型效果评估及无线仿真模型选取策略[J]. 移动通信, 2018, 42(09): 62-67.
- [2] 黄劲安. 基于 Crosswave 射线追踪模型的 TD-LTE 规划仿真[J]. 移动通信, 2014, 38(24): 76-81.
- [3] 徐欣,张厚利. 射线传播(CrossWave)模型变量影响分析[J]. 工程建设与设计, 2016(16): 166-168.
- [4] 马开颜,任心燕. 基于智能手机的成人移动学习[J]. 北京宣武红旗业余大学学报, 2018(02): 68-72.
- [5] 李文洲. 基于智能手机的移动学习在高职英语听说课程中的应用[J]. 课程教育研究, 2018(09): 72-73.
- [6] 汲取在线教育交互式精髓实现高职英语最有效学习—基于在线课程的混合式教学模式的设计与实现[J]. 邢台职业技术学院学报, 2017, 34(06): 2.

参考文献

- [1] 鞠涛. 移动通信系统的资源调度与功率分配技术研究及实现[D]. 东南大学, 2005.
- [2] 徐光勇. 面向节能的4G移动通信网络低能耗功率分配策略[D]. 北京邮电大学, 2013.

(上接第103页)

- [1] 杨辉,杨洪琴. 智能手机移动学习在地方院校教育中的应用研究[J]. 电脑知识与技术, 2018, 14(30): 123-124.
- [2] 朱琪. 基于动画技术的交互式思政教学系统的设计与实现[J]. 自动化与仪器仪表, 2018(09): 107-109.
- [3] 孙佳瑜,王延鹏. 基于移动终端的移动学习在高等数学教学中的应用研究[J]. 黑龙江教育(理论与实践), 2018(04): 49-50.

(上接第118页)左右。

由于站点覆盖范围、场景和厂商设备耗能差异,节电效能的呈现并不完全一致。在低话务低流量区域,节电效果相对更为明显。由于目前整体山西电信的平均设备负荷还处于较低水平,预计在全网开启节能功能后,会产生比较良好的节能效益。