

文章编号: 1000-5641(2017)05-0080-07

在线广告中点击率预测研究

肖 垚¹, 毕军芳², 韩 易¹, 董启文¹

(1. 华东师范大学 数据科学与工程学院, 上海 200062;

2. 长江口水文水资源勘测局, 上海 200136)

摘要: 随着互联网的发展和用户的增长, 广告行业从传统的线下广告模式, 逐步转变为线上广告模式. 同时, 由于大数据分析技术的运用, 线上广告模式相比于传统广告也体现了巨大的优越性. 广告主之间相互竞争, 通过竞价的方式, 将自己的广告投放到运营媒体的广告位上. 所以, 在投放前预测该广告可能被用户点击的概率 (CTR), 对于广告主减少成本和增加可能收入来说非常重要. 本文在调研了目前常用的广告点击率预测模型的基础上, 选取广告主、广告和投放媒体平台信息作为预测模型的特征, 采用真实数据集验证说明各种模型的优劣性, 以及不同特征对广告点击率预测结果的影响.

关键词: 计算广告; CTR; 机器学习

中图分类号: TP391 **文献标志码:** A **DOI:** 10.3969/j.issn.1000-5641.2017.05.008

Study of click through rate prediction in online advertisement

XIAO YAO¹, BI Jun-fang², HAN YI¹, DONG Qi-wen¹

(1. School of Data Science and Engineering, East China Normal University,
Shanghai 200062, China;

2. Yangtze River Estuary Survey Bureau of Hydrology and Water Resource, CWRC,
Ministry of Water Resources, Shanghai 200136, China)

Abstract: With the development of the Internet and the growth of users, the advertising industry originated from the traditional offline advertising model, is gradually transforming into online advertising model. At the same time, due to the use of large data analysis technology, online advertising shows great advantages when compared with traditional advertising. The advertisers deliver their advertisements to the platform's specific positions by competition auction of counterparts. Therefore, it is important to predict the click through rate (CTR) of a given advertisement before auction, which is important for advertisers to reduce costs and expand their likely revenue. This paper introduces the commonly used ad click rate prediction model, uses the information from different

收稿日期: 2017-05-01

基金项目: 国家重点研发计划(2016YFB1000905); 国家自然科学基金广东省联合重点项目(U1401256);
国家自然科学基金(61672234, 61402177); 华东师范大学信息化软课题

第一作者: 肖 垚, 男, 硕士研究生, 研究方向为广告点击率预测.

通信作者: 董启文, 男, 硕士生导师, 副教授, 研究方向为网络信息学、生物信息学.

E-mail: qwdong@dase.ecnu.edu.cn.

advertisers, advertisements and media platforms as the features of machine learning, and uses real data sets to illustrate the advantages of various models, and the impact of different features on the ad click rate.

Key words: computational advertising; CTR; machine learning

0 引言

在线广告起源于20世纪末,当时的媒体网站刚刚起步,随着这些网站的用户数逐渐增多,网站取得了不少的流量规模.投资人希望能将这些流量变现,最初的做法就是把html网站的页面当成传统的杂志版面,将广告插入其中.原来的广告主,也就将这些网站当成一本杂志,按传统线下广告的方式进行广告位的采买.

几十年间的发展,目前在线广告的市场规模达到2 093.7亿元,预计到2018年整体规模有望突破4 000亿元.因此,准确预测广告的点击率,可以提高市场的效率,对于广告主、用户和媒体平台是一件“三赢”的事情.广告主希望更多的人能点击自己的广告,了解自己的产品.媒体平台希望获取更多的利润.用户希望更多合适的广告推送给自己,提升自己的用户体验^[1-2].

对于广告主来说,最大的目标是尽可能用有限的成本,投放最有可能被点击的广告,这样才能获得最大的利益.因此,这就需要提高广告的点击率(CTR).广告点击率的预测需要依赖于广告的历史投放记录(包括点击和未点击的记录).分析广告日志是一个预测和优化的双向过程.通过对点击日志的分析,不但可以预测和优化广告的点击率,还可以优化广告投放页面中广告的排序结果以及估计用户的满意度.因此,点击率预测是一个互联网许多领域都需要解决的问题,包括搜索引擎的排序结果以及推荐系统.广告点击率只是点击率预测的一个应用,但这一应用是全球网络公司的一个重要收入来源,因此有着重要的商业价值和学术研究价值,已经成为了近几年学术界和产业界的一个重要研究领域^[3].

本文从正负样本比例、特征选择和机器学习模型三个方面着手,研究这三点对预测效果的影响,经过一系列的对比实验,选出最合适的预测模型.

1 相关工作介绍

点击率估计是指在给定网页和用户的情况下估计所投放的广告被点击次数占总展示次数的比例.互联网广告的点击率从20世纪90年代起一直呈下降趋势,目前平均点击率在0.2%~0.3%,0.2%的广告点击率即被视为非常成功的广告投放.随着广告计费方式的改变,广告点击率估计在广告投放过程中占有越来越重要的地位,估计的结果直接影响到广告检索结果的排序,进而影响到用户、网络媒体和广告主的效用.据统计所有广告的展示频率和点击率均呈幂率分布^[4].

在广告点击率预测方面,已经有不少成熟的研究.文献[5]运用逻辑回归进行广告点击率的预测,因为逻辑回归的结果在(0-1)区间中和点击概率分布的区间一致.同时这篇文章强调,广告在页面中的位置对于该广告最终点击率的重要性.文献[6]运用增强决策树加逻辑回归进行广告点击率预测.将输入的特征通过增强决策树进行转换,然后将结果作为逻辑回归的输入.基于历史广告数据丰富的预估模型还有贝叶斯模型^[7],决策树模型^[8],支持向量机模型^[9]等.

文献 [10] 采用聚类的方法对点击率进行预测, 根据广告的内容进行聚类. 除了聚类模型, 还有基于因子分解机模型^[11]用了对新广告进行预测. 还有一类采用新的模型对点击数据进行建模, 例如层次结构^[12-13]和时间空间模型^[14].

2 数据集实验方法以及评估指标

2.1 实验数据

为了保证实验结果的严谨性和准确性, 我们使用的实验数据来自真实的企业环境. 广告日志文件分为两种类型: 点击日志和未点击日志, 两种类型日志文件中字段相同. 我们从所有的日志文件中抽取部分数据, 共 60 万条记录进行实验. 实验代码用 python 实现, 实验环境为八核, 64Bit CPU 内存 64G 的 Linux 服务器. 实验采用交叉验证的方式, 将数据分成十份, 每次将其中的九份作为训练数据, 一份作为测试数据, 循环十次. 表 1 代表日志中作为模型预测特征的字段的实际含义.

表 1 日志字段对应的特征

Tab. 1 The log field corresponding to the feature

特征类型	日志字段
一天中的时间段	tis
地域	ip
竞价平台	adx
流量类型	devicetype
平台	platform
浏览器	browser
操作系统	os
广告位ID	adslot_id
广告位位置	adslot_pos
活动ID	campaign_id
活动组ID	group_id
素材ID	creative_id
素材尺寸	creative_size
广告主ID	advertiser_id
广告代理ID	ad_agent_id

2.2 点击率预测模型

工业界使用最多的广告点击率预测模型是逻辑回归模型. 在广告点击率预测这一过程中, 我们将点击事件 h 看成一个二元取值的随机变量, 那么其取值为真 ($h = 1$) 的概率就是点击率, 因此, 点击事件的分布可以写成以点击率 μ 为参数的二项分布 (binomial distribution):

$$P(h) = \mu^h (1 - \mu)^{1-h}. \quad (2.2.1)$$

而点击率预测模型的作用是在 (a, u, c) 组合与点击率 μ 之间建立函数关系, a, u, c 三个变量分别代表广告、用户与环境. 这可以表示成对 $\mu(a, u, c) = P(h = 1|a, u, c)$ 的概率建模问题, 可以很自然地想到逻辑回归:

$$P(h|a, u, c) = \sigma((2h - 1)\omega^T \omega \chi(a, u, c)) = \{1 + e^{-(2h-1)\omega^T \chi(a, u, c)}\}^{-1}. \quad (2.2.2)$$

其中 χ 表示 (a, u, c) 组合上的特征矢量, ω 为各个特征的加权系数, 也就是次模型需要优化参数; $(2h-1)\omega^T\chi$ 这一线性函数的输出经过 sigmoid 函数映射到 $(0,1)$ 区间内, 其中 $(2h-1)$ 是为了将 $\{0,1\}$ 的点击变量变换到集合 $\{-1,1\}$ 上. 显然, 广告的点击率与广告内容、用户和上下文环境三个因素都有关系^[5].

决策树也是一类常见的点击率预测模型, 对于二分类问题来说, 我们希望从给定训练数据集学得一个模型用来对新样本进行分类. 比如, 对于广告点击率预测来说, 我们可以将其看成对“当前广告是否会被点击”这个问题的决策过程. 通常, 一颗决策树包含一个根结点, 还有若干个内部节点和叶子节点. 根结点包含的是样本的全集, 叶子节点对应最终的决策结果. 因此, 该模型训练中, 就是通过计算每个节点划分后样本的纯度, 来形成一个完整决策树. 计算样本的纯度, 通常有三种评价标准: 信息熵、增益率和基尼系数. 通过三种评价标准构造的决策树依次称为 ID3 决策树、C4.5 决策树和 CART 决策树. 本实验中使用的是 CART 决策树, 因为 CART 决策树能够处理数据型和类别型的属性.

随机森林 (Random Forest) 属于一种集成学习方法, 它是对决策树的一种扩展, 通过随机构造多棵决策树, 最终通过投票方式决定预测结果. 当测试数据进入随机森林时, 其实就是让每一棵决策树进行分类, 所有数据的分类结果按照决策树投票多少的分数而定. 因此随机森林是一个包含多个决策树的分类器, 并且其输出的类别是由个别树输出的类别的众数而定.

梯度提升决策树 (Gradient Boosting Descent Tree)^[15], 是一种迭代的决策树算法, 该算法由多棵决策树组成, 所有树的分类结果累加起来作为最终答案. 它在被提出之初就和支持向量机 (SVM) 一起被认为是泛化能力较强的算法. 近年来更因为被用于搜索排序领域而引起广泛关注^[16].

2.3 评估指标

为了检查模型的训练效果, 我们采用三个评价指标: 准确率 (Precision)、ROC 曲线下面积 AUC^[17]、对数损失 (Logloss).

准确率: 也就是对于训练数据中的 N 条数据, 统计系统能够判断准确的数据条数 M , 最后进行简单相除得到 M/N 作为评判标准.

准确率确实是一个很好很直观的评价指标, 但是有时候准确率高并不能代表一个模型就好. 比如某个地区某天地震的预测, 假设我们有一组特征作为地震分类的属性, 预测结果只有两种: 不发生地震和发生地震 (记为 0 和 1). 由于在日常生活中, 地震的发生是一个小概率事件, 在一个分类模型中, 如果将每一个测试用例类别都划分为 0, 那么它就可能达到 99% 的正确率, 但是这个模型毫无意义. 所以, 准确率并不能单独作为判断模型性能好坏的指标.

因此我们引用 AUC 和 Logloss 这两个指标来评判模型的好坏. ROC 是受试者工作特征曲线 (receiver operating characteristic curve), 又称为感受性曲线 (sensitivity curve). 对于上面的准确率指标来说, 评价分类结果为正类或负类的阈值为 0.5 (模型训练的结果大于 0.5 时判别为正类, 模型训练的结果小于 0.5 时判别为负类). 而 ROC 曲线是利用一系列的阈值作为评判正负类的依据. AUC 就是 ROC 曲线与 x 轴围成的面积. 当 AUC 的值越大, 模型的效果意味着也越好. 对数损失函数:

$$\text{logloss} = -\log P(y_t|y_p) = -(y_t \log(y_p) + (1 - y_t) \log(1 - y_p)). \quad (2.3.1)$$

y_t 为真实值, y_p 为预测值. 由上述公式可以看出, 当预测值与真实值相等时结果为 0, 此时预测百分百正确, 因此结果越小则表明模型的预测能力越好.

3 实验结果

3.1 正负样本比例对模型预测的影响

在广告点击率预测场景下, 正负样本比例不均衡是普遍存在的问题. 本组实验目的是研究正负样本比例对模型预测性能的影响. 从日志文件中, 依次按比例随机抽取正负样本数据, 作为训练数据. 通过这组对比实验, 我们可以得出最佳的正负样本比例, 应用于广告点击率预测中.

实验的结果如图 1 所示, 从图中我们可以看出, 正负样本的比例相差过大的时候, 模型的效果并不是理想的, 最理想的效果应该是正负样本的比例接近. 可以认为, 当正负样本比例相差过大时, 模型会把比例少的样本数据误当成噪音数据, 从而影响了模型预测的效果.

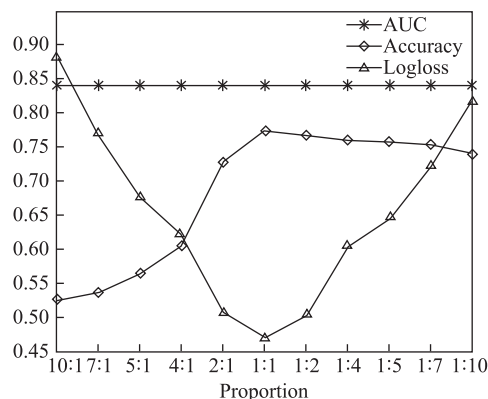


图 1 不同正负样本比例对模型的影响

Fig. 1 Influence of Different Positive and Negative Sample Proportions on Model

我们选取的正负样本比例依次从 10:1 到 1:10, 从结果可以看出, 对于 Precision, 和 Logloss 来说, 当正负样本比例为 1:1 时, 效果最好, Precision 将近 0.77, Logloss 低至 0.47. 当采取不同的正负样本比例时, Precision 和 Logloss 值关于 1:1 的结果对称, 例如 10:1 和 1:10 的结果很相似.

对于 AUC 来说, 不同的比例, 变化不大, 主要是 AUC 利用不同的阈值来作为判断正负类的依据. 计算 AUC 值时, 会依次将预测样本为正样本的概率作为判断是否为点击的阈值. 当测试样本属于正样本的概率大于或等于这个阈值时, 我们认为它为正确样本, 否则为负样本. 因此选取不同的样本比例时, 可能会得到相似的预测样本为正样本的概率, 最终得到的 AUC 值也会相近.

3.2 不同特征对模型预测的影响

在广告的点击和未点击的历史日志中, 包含有很多的字段比如, 广告的编号, 广告在网页的位置, 用户的浏览时间等, 这些字段都能作为预测 CTR 的特征, 但是并不是特征越多效果越理想. 特征的选取关键在于正确而非数量多, 文献 [6] 指出模型效果的关键在于找到最大价值的特征, 一旦选择的特征和模型是正确的, 那么其余因素对模型的效果影响是甚微的. 同时, 如果使用的特征维数过多, 也会影响模型训练的速度, 增加了内存和 CPU 的消耗, 也抑制了模型的性能. 本文中, 先选取广告信息的特征作为基础实验, 然后逐渐增加用户的信息和媒体平台的信息作为特征, 对比模型性能的变化. 本节中的机器学习模型使用的仍然是逻辑回归.

文献 [18] 指出, 广告的点击率与广告的质量有关. 不同广告的质量是不同的, 如表 2 所示, 我们可以通过训练数据中的 adslot_id(广告位 ID)字段来唯一标识一个广告. 文献 [19] 指出, 广告的点击率会随着广告在一个网站中的位置排名变化而变化, 这是因为位置排名低的广告不会被人们注意到. 文献 [20] 也证实了, 广告出现在网页中的位置会对用户造成信任偏差, 影响人

们的点击情况. 广告的位置, 在实验数据中有 `adslot_pos`(广告位位置)这个字段来表示, 我们以 `adslot_id`(广告位 ID)和 `adslot_pos`(广告位位置)和 `creative_id`(素材 ID)等作为广告的特征. 然后再逐渐增加用户的特征如用户点击时间 `tis`, 用户点击的设备 `device_type`, 用户地址 `ip` 等和上下文环境的特征, 比如广告显示的媒体平台 `adx`.

表 2 不同特征对模型预测结果的影响

Tab. 2 Impact of Different Features on Model Prediction

特征 评价指标	广告信息	广告信息+用户信息	广告信息+用户信息+媒体平台信息
Precision	0.705 4	0.764 4	0.774 4
Auc	0.704 9	0.830 3	0.836 9
Logloss	0.598 5	0.480 8	0.474 9

从实验结果看出, 仅用广告信息作为特征, 能取得一定好的效果, 准确率达到了 0.70. 当加上用户的信息后, 预测的效果能得到比较大的提升. Precision 方面提高了 0.06 个百分点, 这对于广告点击率预测来说已经是很大的提升. Logloss 进一步降低了 0.11, Auc 提高了 0.13, 由此看出, 当增加用户信息后, 模型的性能提升很明显.

但是, 最后加入媒体平台信息特征后, 模型的性能提升效果不明显, 三个评价指标来看, 模型性能仅有很微小的提升. 原因是, 我们所拥有的关于媒体平台的信息太少, 不能完全涵盖媒体平台的相关信息. 只有表明平台名称 (`adx`) 的这一个字段. 如果日记中这部分信息能完善, 那么模型的预测效果能得到更多的提升.

总之, 对于广告点击率预测, 广告本身的信息对于模型预测最有帮助, 除了本身信息部分, 用户的信息和平台的信息也能有效地增加预测的准确性.

3.3 不同机器学习模型对比

从图 2 可以看出逻辑回归、决策树、随机森林、梯度提升决策树这四个模型的性能逐渐提升. 其中梯度提升决策树性能最好, 准确率能达到 0.880, 对数损失低至 0.4, AUC 接近 0.85. 与原始的逻辑回归相比, 性能提高了很多. 决策树与逻辑回归相比, 仅有微弱的提升, 但是通过构造多颗决策树形成一个随机森林时, 性能提高很明显. 但随机森林与梯度提升决策树相比, 还是有微弱的劣势.

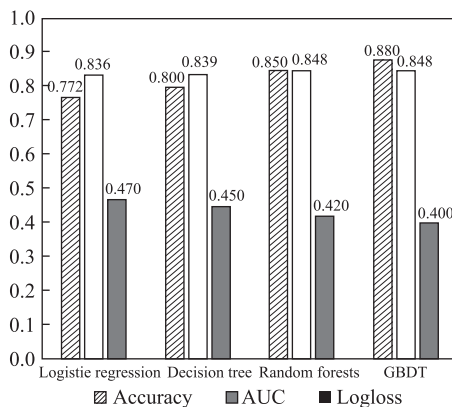


图 2 不同模型预测性能的对比

Fig. 2 Comparison of Predictive Performance of Different Models

虽然这四种模型的性能依次提升, 但是训练消耗的时间也在依次增长, 相同的数据量, 梯度提升决策树训练的时间大概是逻辑回归训练时间的两倍左右. 训练使用的 60 万条数据, 使用梯

度提升决策树训练所需时间超过 20 min, 而使用随机森林, 大约在 15 min, 逻辑回归和决策树使用的时间比较少, 大概 10 min 左右.

逻辑回归虽然效果一般, 却胜在模型的可解释性强, 它拟合出来的参数就代表了每一个特征对结果的影响. 也是一个理解数据的好工具. 决策树能够生成清晰的基于特征选择不同预测结果的树状结构, 随机森林在现实分析中被大量使用, 它相对于决策树, 在准确性上有了很大的提升. 梯度提升决策树同随机森林一样, 不容易陷入过拟合, 而且能够得到很高的精度.

4 结 论

本文通过一系列实验研究了广告点击率的预测问题. 对于训练数据来说, 要选择合适的正负样本比例. 训练数据中点击和未点击广告数据的比例会影响模型预测的准确率. 实验结果表明, 当正负样本比例接近的时候, 模型能取得好的预测效果. 当正负样本比例不均衡的时候, 会在预测中产生很大的错误率.

特征的选择对于模型的预测性能也有至关重要的影响. 广告点击率预测中, 能运用的特征有很多. 特征包含广告本身、用户和媒体平台三者的信息. 模型要达到一个好的准确率, 那么就要充分挖掘这三者的信息作为特征. 当使用的三种类型的特征组合得越好, 模型性能就能够越好. 实验中, 依次增加广告特征、用户特征和媒体平台特征, 从结果中可以看出, 所含的广告、用户和网站上下文信息越充分的时候, 模型的预测效果越佳.

逻辑回归、决策树、随机森林和梯度提升决策树这四个广告点击率预测常用的模型中, 梯度提升决策树和随机森林的预测效果较好, 不过, 模型训练需要的时间较长. 逻辑回归和决策树的性能虽然劣于梯度提升决策树和随机森林, 但是其训练时间较短.

[参 考 文 献]

- [1] GABRILOVICH E. An Overview of Computational Advertising[R/OL]. [2013-03-21]. <http://research.yahoo.com/pub/2915>.
- [2] AGARWAL D, CHAKRABARTI D. Statistical Challenge in Online Advertising[R/OL]. [2013-03-21]. <http://research.yahoo.com/pub/2430>.
- [3] 纪文迪, 王晓玲, 周傲英. 广告点击率估算技术综述[J]. 华东师范大学学报(自然科学版), 2013(3): 2-14.
- [4] AGARWAL D, AGRAWAL R, KHANNA R, et al. Estimating rates of rare events with multiple hierarchies through scalable log-linear models[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2010: 213-222.
- [5] RICHARDSON M, DOMINOWSKA E, RAGNO R. Predicting clicks:estimating the click-through rate for new ads[C]// International Conference on World Wide Web. ACM, 2007: 521-530.
- [6] HE X, PAN J, JIN O, et al. Practical Lessons from Predicting Clicks on Ads at Facebook[C]// Eighth International Workshop on Data Mining for Online Advertising. ACM, 2014: 1-9.
- [7] CHAPELLE O, ZHANG Y. A dynamic bayesian network click model for web search ranking[C]// International Conference on World Wide Web. ACM, 2009: 1-10.
- [8] DUPRET G E, PIWOWARSKI B. A user browsing model to predict search engine click data from past observations[C]// International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2008: 331-338.
- [9] DAVE K, VARMA V. Predicting the click-through rate for rare/new ads[R]. Center for Search and Information Extraction Lab International Institute of Information Technology Hyderabad, INDIA, 2010.
- [10] REGELSON M, FAIN D. Predicting click-through rate using keyword clusters[C]//Proceedings of the Second Workshop on Sponsored Search Auctions, 2006: 9623.
- [11] RENDLE S. Factorization machines[C]// IEEE International Conference on Data Mining. IEEE Computer Society, 2010: 995-1000.

(下转第 100 页)

- [18] RASTAGI V, NATH S. Differentially private aggregation of distributed time-series with transformation and encryption[C]// Proc of the Management of data. Indiana: ACM, 2010: 6-11
- [19] SARATHY R, MURALIDHAR K. Evaluating laplace noise addition to satisfy differential privacy for numeric data[J]. Transactions on Data Privacy, 2011, 4(1): 1-17.
- [20] SHAMIR A. How to share a secret[J]. Communications of the ACM, 1979, 22(11): 612-613.
- [21] TIAN X X, SHA C F, WANG X L, et al. Privacy preserving query processing on secret share based data storage[C]// Proc of the Database Systems for Advanced Applications. Hong Kong: Springer, 2011: 108-122.
- [22] RASTOGI V, NATH S. Differentially private aggregation of distributed time-series with transformation and encryption[R]. Tech Rep MSR-TR-2009-186, Microsoft Research, 2009.
- [23] LI Q D, ZHOU Y H. Research and application based on A. Shamir's (t, n) threshold secret sharing scheme[C]// Proc of the Computer Science & Education. Melbourne: IEEE, 2012(6): 14-17.
- [24] DWORK C, MCSHERRY F, NISSIM K, et al. Calibrating noise to sensitivity in private data analysis[C]// Proc of the 3rd Theory of Cryptography Conference. New York: Springer, 2006: 265-284.
- [25] DWORK C. Differential privacy: A survey of results[C]// Proc of the Theory and Applications of Models of Computation. China: Springer, 2008: 1-19.
- [26] 田秀霞, 高明, 王晓玲, 等. 数据库服务-安全与隐私保护 [J]. 软件学报, 2010, 21(5): 991-1006.
- [27] CHAUM D. Blind signatures for untraceable payments[C]// Proc of the Advances in Cryptology. USA: Springer, 1982: 199-203.
- [28] 张明武, 杨波, 祝胜林. 可信模块隐私保护的自证明签名方案 [J]. 北京邮电大学学报, 2009, 32(1): 60-64.
- [29] LIUY L, JIN Z G. Security enhancement of WAPI access authentication protocol(WAI)[J]. Journal of Harbin Institute of Tehnolo (New Series), 2012, 19(6): 42-46.

(责任编辑: 张 晶)

(上接第 86 页)

- [12] WANG X, LI W, CUI Y, et al. Click-through rate estimation for rare events in online advertising[G]//HUA X S, MEI T, HANJALIC A. Online Multimedia Advertising: Techniques and Technologies . Hershey: IGI Global, 2010. doi: 10.4018/978-1-60960-189-8.ch001.
- [13] AGARWAL D, BRODER A Z, CHAKRABARTI D, et al. Estimating rates of rare events at multiple resolutions[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - Kdd. ACM, 2007: 16-25.
- [14] AGARWAL D, CHEN B C, ELANGO P. Spatio-temporal models for estimating click-through rate[C]// International Conference on World Wide Web. ACM, 2009: 21-30.
- [15] SCHONLAU M. Boosted regression (boosting): An introductory tutorial and a stata plugin[J]. Stata Journal, 2005, 5(3): 330-354.
- [16] BURGESS C J C. From ranknet to lambdarank to lambdamart: An overview[R]. Microsoft Research Technical Report, 2010.
- [17] FANG Y, LIU J. A novel prior-based real-time click through rate prediction model[J]. International Journal of Machine Learning & Cybernetics, 2014, 5(6): 887-895.
- [18] FAIR D C, PEDERSEN J O. Sponsored search: A brief history[J]. Bulletin of the American Society for Information Science & Technology, 2010, 32(2): 12-13.
- [19] RICHARDSON M, DOMINOWSKA E, RAGNO R. Predicting clicks:estimating the click-through rate for new ads[C]// International Conference on World Wide Web. ACM, 2007: 521-530.
- [20] JOACHIMS T, GRANKA L, PAN B, et al. Accurately interpreting clickthrough data as implicit feedback[C]// Proceedings of the 28th Annual International ACM SIGIR, 2005: 154-161.

(责任编辑: 李万会)