

硕士研究生学位论文

|  |  |
| --- | --- |
| 题目： | 面向微博的在线广告点击 |
|  | 行为预测系统设计与实现 |

|  |  |
| --- | --- |
| 姓 名： | 孙超 |
| 学 号： | 1301221193 |
| 院 系： | 软件与微电子学院 |
| 专 业： | 软件工程 |
| 研究方向： | 互联网信息挖掘与处理 |
| 导师姓名： | 王海峰、刘宏志 |

二〇一 年 月

版权声明

任何收存和保管本论文各种版本的单位和个人，未经本论文作者同意，不得将本论文转借他人，亦不得随意复制、抄录、拍照或以任何方式传播。否则，引起有碍作者著作权之问题，将可能承担法律责任。

# 摘要

广告点击行为预测是互联网计算广告的重要环节，通常的做法是先用机器学习算法在已有用户的特征数据和点击行为记录上训练出一个模型，然后用训练出的模型结合新的用户特征数据返回该用户点击行为的概率。以往工作的重点是训练模型参数选取和用户特征选取上，其中用户特征选取多集中于结构化特征，即如用户的性别、年龄、职业、教育程度等固有属性。

本文在用户结构化特征基础上，利用用户的微博行为数据（包括发表、转发、点赞）过的文章，通过特征工程转换为用户的非结构化特征，用GBDT和LR融合的方法做点击行为预估，并设计和实现了一个可商业化运营的线上分布式系统。实验和线上运行记录表明该方法较传统的只用结构化特征的方法在预估准确率方面有了明显的提高。

关键词：广告，点击率，预测

ENGLISH TITLE

Author Name ( Major )

Directed by your Supervisor

# ABSTRACT

In environmental economics, environmental resources including environmental quality are categorized as amenity resources. Due to its importance to human welfare, the amenity resources theoretical study and valuation is an ongoing issue at the academic frontier in the environmental economics area.

KEY WORDS: Key word 1, Key word 2, Key word 3, ……

# 目录

[摘要 I](#_Toc469005005)

[ABSTRACT II](#_Toc469005006)

[目录 III](#_Toc469005007)

[第一章 绪论 1](#_Toc469005008)

[1.1 项目背景 1](#_Toc469005009)

[1.1.1 Ⅲ族氮化物半导体的晶体结构 1](#_Toc469005010)

[第二章 研究进展 2](#_Toc469005011)

[2.1 环境中黑炭的主要来源 2](#_Toc469005012)

[第四章 图表示例 4](#_Toc469005013)

[第五章 结论及展望 5](#_Toc469005014)

[参考文献 6](#_Toc469005015)

[附录A 附录示例 8](#_Toc469005016)

[致谢 9](#_Toc469005017)

[北京大学学位论文原创性声明和使用授权说明 10](#_Toc469005018)

注：目录从第1章开始，前边因页眉需要设置了标题，实际使用时更新后去掉前边部分。使用时请删除本注释。如本示例，更新目录后删除前边三项（摘要、ABSTRACT、目录）即可。

# 第一章 绪论

## 1.1 项目背景

当前我国互联网产业迅速发展，根据中国互联网络信息中心（CNNIC）的最新统计，截至2016年6月，我国网民规模达7.10亿，互联网普及率达到51.7%，与2015年底相比提高1.3个百分点，超过全球平均水平3.1个百分点，超过亚洲平均水平8.1个百分点[1]。与此同时，电子商务产业也发展迅速，十二五期间，中国的网络零售交易额规模跃居世界第一，网购在网民中的普及率高达55.7%，仅在2016年双11期间，天猫交易额就突破1200亿元，物流订单量超6.5亿[2]。

在此背景下催生了互联网广告这一新兴产业，进而催生了计算广告学这一新兴学科。相比于基于传统媒体如电视、广播、报纸海量而无特定目标的广告投放，以互联网为媒介的广告投放可以根据用户特征（主要包括用户个人信息和上网行为）做到个性化精准投放。从而为大大提高了广告收益，为企业节省了推广成本。其中，广告点击行为预测是计算广告学的一个重要研究内容，目前主要方法是使用机器学习算法在用户特征数据和行为数据上训练得到一个模型，再用训练得到的模型在线预测新用户点击广告的概率。

### 1.1.1 Ⅲ族氮化物半导体的晶体结构

………….

# 第二章 研究进展[[1]](#footnote-1)

## 2.1 环境中黑炭的主要来源

环境中黑炭 (black carbon)气溶胶的主要来源包括各种化石燃料和生物质燃料的不完全燃烧过程 (Penner et al., 1993; Bond et al., 2004)，这些不完全燃烧在自然界和人类活动中都会发生，因此，环境中黑炭气溶胶的来源十分广泛。对当今大气环境中的黑炭，其主要来源是人类相关的燃料燃烧活动 (段凤魁,2007)，此外，一些自然过程也会产生黑炭，如森林火灾、草原火灾等。根据过去的排放清单研究，大气环境中黑炭气溶胶的来源主要包括：1) 有机燃料的燃烧，主要包括能源行业、工业部门、交通运输行业、居民生活中煤、石油、天然气和各种生物质燃料的使用。通常而言，燃烧效率越高，产生的黑炭气溶胶的量越低；2) 工业炼焦，主要包括炼焦过程中的炼制过程、焦炉加热系统以及焦炉煤气的泄漏等等；3) 工业制砖，主要包括制砖过程中物料破碎输送、坯体人工干燥和烧成工段等过程；4) 垃圾焚烧，包括生活垃圾和工业废料的燃烧过程；5) 天然火灾和野外农业废弃物燃烧，包括森林、草原火灾和秸秆的燃烧。目前大部分研究表明，民用取暖和做饭过程中的燃料燃烧和城市柴油车是黑炭气溶胶大气排放量最大的源 (Streets et al., 2001, 2003, 2013; Bond et al., 2004, 2006; Cao et al., 2006; Klimont et al., 2009; Zhang et al., 2009; Lu et al., 2011)。

……

# 第四章 图表示例[[2]](#footnote-2)



图3.15 全球NAT-CO2-2007清单与PKU-CO2-2007比较的空间示意图

表3.5 室外细菌气溶胶香农-维纳指数（H）和均匀性指数（E）

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Stage 1 (>7.1 μm)** | | | | **Stage 2 (4.8-7.1 μm)** | | | | **Stage 3 (3.2-4.7 μm)** | | | |
| Con | Low | Medium | High | Con | Low | Medium | High | Con | Low | Medium | High |
| **H** | **2.52** | 2.58 | 2.57 | ***2.24*** | **2.48** | 2.21 | 2.21 | ***2.36*** | **2.66** | 2.65 | 2.64 | ***2.53*** |
| **E** | 0.87 | 0.88 | 0.93 | 0.85 | 0.9 | 0.86 | 0.86 | 0.85 | 0.9 | 0.9 | 0.85 | 0.88 |

# 第五章 结论及展望

# 参考文献[[3]](#footnote-3)

1. 第38次中国互联网络发展状况统计报告 <http://www.cnnic.net.cn/>
2. 新浪财经频道 <http://finance.sina.com.cn/roll/2016-11-12/doc-ifxxsfip4572557.shtml>
3. 刘鹏, 王超, 计算广告 互联网商业变现的市场与技术, 人民邮电出版社, 2015

[1] Intel Corperation, http://download.intel.com/pressroom/kits/IntelProcessorHistory.pdf.

[2] Intel Corperation, http://www.intel.com/content/dam/www/public/us/en/documents/presentation/  
revolutionary-22nm-transistor-technology-presentation.pdf.

[3] I. Žutić, J. Fabian and S. Das Sarma, Spintronics: Fundamentals and applications, Reviews of Modern Physics 76 (2), 323-410 (2004).

[4] R. Hanson, L. Kouwenhoven, J. Petta, S. Tarucha and L. Vandersypen, Spins in few-electron quantum dots, Reviews of Modern Physics 79 (4), 1217 (2007).

[5] D. Loss and D. P. DiVincenzo, Quantum computation with quantum dots,Physical Review A 57 (1), 120 (1998).

注：以上是“顺序编码制”索引文献时参考文献著录法（对应第1章示例）。各项著录信息未核准，仅为样式参考。“著者—出版年”制索引文献著录方法如下（对应第二章示例）：

段凤魁, 贺克斌, 刘咸德, 董树屏, 杨复沫. 2007. 含碳气溶胶研究进展：有机碳和元素碳. 环境工程学报, 1: 1-8.

Bond T.C.; Bergstrom R.W. 2006. Light absorption by carbonaceous particles: an investigative review. Aerosol Science and Technology, 40: 27-67.

Bond, T. C.; Streets, D. G.; Yarber, K. F.; et al. 2004. A technology-based global inventory of black and organic carbon emissions from combustion. Journal of Geophysical Research, 109, D14203.

Cao, G. L.; Zhang, X. Y.; Zheng, F. C. 2006. Inventory of black carbon and organic carbon emissions from China. Atmospheric Environment, 40: 6516-6527.

Klimont, Z.; et al. 2009. Projections of SO2, NOx and carbonaceous aerosols emissions in Asia. Tellus, 61B, 602-617.

Lu, Z.; Zhang, Q.; Streets, D. G. 2011. Sulfur dioxide and primary carbonaceous aerosol emissions in China and India, 1996-2010. Atmospheric Chemistry and Physics, 11, 9839-9864.

Penner, J. E.; Eddleman, H.; Novakov, T. 1993. Towards the development of a global inventory for black carbon emissions. Atmospheric Environment, 27 (A): 1277-1295.

Streets, D. G.; Bond, T. C.; Carmichael, G. R.; et al. 2003. An inventory of gaseous and primary aerosol emissions in Asia in the year 2000. Journal of Geophysical Research, 108, 8809.

Streets, D. G.; Bond, T. C.; Lee, T.; Jang, C. 2004. On the future of carbonaceous aerosol emissions. Journal of Geophysical Research, 109, D24212, doi:10.1029/2004JD004902.

Streets, D.G.; Shalini, G.; Waldhoff, S.T.; et al. 2001. Michae Black carbon emissions in China. Atmospheric Environment, 35, 4281- 4296.

Zhang, Q.; Streets, D. G.; Carmichael, G. R.; et al. 2009. Asian emissions in 2006 for the NASA INTEX-B mission. Atmospheric Chemistry Physics, 9, 4081- 4139.

# 附录A 附录示例

# 致谢

本论文是在xx老师的悉心指导下完成的。xx老师作为一名优秀的、经验丰富的教师，具有丰富的xx知识和xx经验，在整个论文实验和论文写作过程中，对我进行了耐心的指导和帮助，提出严格要求，引导我不断开阔思路，为我答疑解惑，鼓励我大胆创新，使我在这一段宝贵的时光中，既增长了知识、开阔了视野、锻炼了心态，又培养了良好的实验习惯和科研精神。在此，我向我的指导老师表示最诚挚的谢意！

……

(仅为网络示例，可根据论文实际进行撰写，使用时把模板示例内容尽皆删除即可)

# 北京大学学位论文原创性声明和使用授权说明

**原创性声明**

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品或成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

论文作者签名： 日期： 年 月 日

**学位论文使用授权说明**

（必须装订在提交学校图书馆的印刷本）

本人完全了解北京大学关于收集、保存、使用学位论文的规定，即：

* 按照学校要求提交学位论文的印刷本和电子版本；
* 学校有权保存学位论文的印刷本和电子版，并提供目录检索与阅览服务，在校园网上提供服务；
* 学校可以采用影印、缩印、数字化或其它复制手段保存论文；
* 因某种特殊原因需要延迟发布学位论文电子版，授权学校□一年/□两年/□三年以后，在校园网上全文发布。

（保密论文在解密后遵守此规定）

论文作者签名： 导师签名：

日期： 年 月 日

1. 本章为“著者-出版年制”索引文献示例，实际写作时只能选择本章和第1章索引文献方法之一，不得混用。 [↑](#footnote-ref-1)
2. 图标题在图下方，表标题在表上方。图表序号分章设置，如图3.15表示第三章第15幅图。 [↑](#footnote-ref-2)
3. 全文参考文献索引方式只能选用“顺序编码制”或“著者—出版年制”其中之一，文献列表也应选择相对应的著录方法，此处作为示例列举了两种方式，实际撰写论文时不得混用。 [↑](#footnote-ref-3)