Mining of Massive Datasets

Mining of Massive Datasets_下载链接1_

著者:Jure Leskovec

出版者:Cambridge University Press

出版时间:2014-12-29

装帧:Hardcover

isbn:9781107077232

Written by leading authorities in database and Web technologies, this book is essential reading for students and practitioners alike. The popularity of the Web and Internet commerce provides many extremely large datasets from which information can be gleaned by data mining. This book focuses on practical algorithms that have been used to solve key problems in data mining and can be applied successfully to even the largest datasets. It begins with a discussion of the map-reduce framework, an important tool for parallelizing algorithms automatically. The authors explain the tricks of locality-sensitive hashing and stream processing algorithms for mining data that arrives too fast for exhaustive processing. Other chapters cover the PageRank idea and related tricks for organizing the Web, the problems of finding frequent itemsets and clustering. This second edition includes new and extended coverage on social networks, machine learning and dimensionality reduction.

作者介绍:

Jure Leskovec is Assistant Professor of Computer Science at Stanford University. His research focuses on mining large social and information networks. Problems he

investigates are motivated by large scale data, the Web and on-line media. This research has won several awards including a Microsoft Research Faculty Fellowship, the Alfred P. Sloan Fellowship, Okawa Foundation Fellowship, and numerous best paper awards. His research has also been featured in popular press outlets such as the New York Times, the Wall Street Journal, the Washington Post, MIT Technology Review, NBC, BBC, CBC and Wired. Leskovec has also authored the Stanford Network Analysis Platform (SNAP, http://snap.stanford.edu), a general purpose network analysis and graph mining library that easily scales to massive networks with hundreds of millions of nodes and billions of edges. You can follow him on Twitter at @jure.

目录:

Mining of Massive Datasets_下载链接1_

标签

数据挖掘

计算机

机器学习

Data

Coursera

CS

数据分析

软件工程

评论

bug非常之多, 还找不到地方提交, 读起来极度痛苦, 前看后忘,

也许里面的算法本质上就是这样, bottom line至少近15年最新的论文成果被这么串讲一下, 本科生也能看懂

勉强一刷吧。到时配合斯坦福的课再过一遍~

行文很流畅，看到下面很多人说翻译的问题，由此推荐原版。配合网课还是挺浅显的，例子举得也挺多，自学也可以。步骤写的也很细，有条件完全可以照着码，不晦涩，小白很喜欢。

下学期课程参考textbook，听说professor还不错，打算好好学一下这门课

内容不错，但作为技术向的书有些浮于表面。

花费6个月时间，断断续续看完，哈希和近似的想法真是开阔了眼界。第一回看比较急促，此书值得反复看，多实践。

Mining of Massive Datasets_下载链接1_

书评

看到好多人说这本书是大纲，是目录，没啥内容，讲的浅。 那就对了。
本书是Stanford CS246课程MMDS使用的讲义，还有配套的Slides和HW，所以观看本书请配套课程进行学习，同时coursera上也有配套的课程。 See more detail: http://www.mmds.org/

从总体安排来看，书的结构还是不错的。没看过英文的，但是中文版的行文真的不好，磕磕绊绊看了一半以后实在是没有兴趣看后面的了。
之前了解的pagerank看了以后了解了，之前不了解的adwords还是不了解，

麻烦支那猪以后翻译外文书籍，先找个稍微懂行的把书看一遍行吗！
鉴于中文翻译缩水不准的情况，本掉千辛万苦找来英文原版，一看到目录，本屌就硬了，尼玛作者太牛逼了！
最新补充一句，话说如果这本书的名字叫做类似《数据挖掘基础》的话，本屌绝壁不喷它。本来就是基础的基...

我真的不能忍受一帮子没读过此书，没写过代码，没搞过大数据的外行人在这边乱喷这本书。对豆瓣这本书的评价实在是太失望了。
这是我读到的第一本真正讲“大数据”思路的书。
面对海量数据的时候，我们的软件架构也会跟着发生变化。当你的数据量在内存里放不下的时候，你就得考...

读技术书于我而言就像高中物理老师说的那样：一看就懂、一说就糊、一写就错。为了不马上遗忘昨天刚刚看完的这本书，决定写点东西以帮助多少年之后还有那么一点点记忆。好吧，开写。 1.
总体来说，数据挖掘时数据模型的发现过程。而数据建模的方法可以归纳为两种：数...

很差是给中译版的。
本书的中译版是中科院计算所的王斌老师翻译的，但是翻译的很屎。估计王老师拿到英文稿之后就扔给学生去翻译了，看这翻译水平，实在是不敢恭维。
以上纯为发泄心中不满所写。因为我看译者序，说是自己独立翻译，前后持续了七个多月，并历经多次修改。如果...

只看了两章，所有真心不好打分。这其实是本数学书，而且是一本入门书。这本书的目标读者不是工程师，而是读研或者读博的学生。如果你本身就有数据挖掘后者机器学习的背景，或者就是很喜欢数学，我还是很推荐这本书的，学习新东西总是很有趣的。

看有同学说是
stanford的入门课程，按理说应该不是太难。作为初学者来说，本书翻译的实在不敢恭维，看了50多页是一头雾水，很多话实在是晦涩难懂。本书作用入门级课程来说，基本上涵盖了数据挖掘的各个大类，如果想细致研究某个领域的大拿就不用看了

内容是算法分析应该有的套路, 对于Correctness, Running Time, Storage的证明;
讲得很细, 一个星期要讲3个算法, 看懂以后全部忘光大概率要发生.
要是能多给些直觉解释就好了. Ullman的表达绝对是有问题的, 谁不承认谁就是不客观,
常常一句话我要琢磨2个小时, 比如DGIM算法有一...

本来是计划读英文版《Mining of Massive
Datasets》的，但看到打折，而且译者在序言中信誓旦旦地说翻译的很用心，就买了中文的。结果读了第一章就读不下去了，中文表述太烂了，很多句子让人产生无限歧义，磕磕绊绊，叫人生厌。因此决定再次放弃这样的中文翻译书。

Web数据挖掘特点，相比较ML增加了哪些理论和技术？ （1）
大约覆盖了20篇论文。用了统一的语言，统一深度数学来表达。 （2）
Hash用的特别多。方式各异。如下。 a. 提高检索速度，如index b. 数据随机分组。 c.
定义数据映射，重复这些映射。最基本功能。但对于新数据映射会存...

并非传统的”数据挖掘”教材，更像是，“数据挖掘”在互联网的应用场景，所遇到的问题（数据量大）和解决方案； 不过老实说，这本书挺不好懂的。 大概 get
了几个不错的思想：
思想-1：务必充分利用数据的”稀疏性”， 如数据充分稀疏时，可以利用 HASH
将数据“聚合”成“有效...

这本书其实挺好的，但是真得看英文版。
这是我们上课的参考书之一，英文版有的地方没看懂，就打算找个中文版来看。看了中文版发现，这个翻译的水平基本是跟我大四，研一给老师翻译文章的水平一样的，可以看出这本书应该是找学生翻译的，而且是对专业领域还了解不深的学生翻译的...

看到开篇的两个例子，一个是地图聚类分析伦敦病毒问题，另一个是概率统计的例子。对本书还挺有期望。结果翻到第三章开始，这。。
尼玛整本书就是个目录啊。全书结构如下：知识点，摘要，奇葩的例子，习题。
然后另一个知识点，知识点，识点。。 如果为了平时聊天增加些谈资偶...

终于看完了这本书，读的比较粗，但是还是发现了很多的小错误，不知道是作者的错误

还是译者的错误，总之给人不严谨不严肃的印象，知识还是比较容易理解的（虽然本人没记住多少。。汗。。），还是积累了不错的知识，天道酬勤！

当今时代大规模数据爆炸的速度是惊人的，当然，其应用也是越来越广泛的，从传统的零售业到复杂的商业世界，到处都能见到它的身影。那么大数据有什么典型特征呢？即数据类型繁多、数据体量巨大、价值密度低即处理速度快。本书也正是将注意力集中在了极大规模数据上的挖掘，而且...

Mining of Massive Datasets_下载链接1_