资料

机器学习中更多的数据总是优于更好的算法吗？

<http://mp.weixin.qq.com/s?__biz=MzA4NjA4MTkzMw==&mid=209956288&idx=2&sn=9ab700bf4fa06c4811c3abab578c2e06&scene=1&srcid=0929kYV8CtZC60vkPN0y9cwa#rd>

Halevy A, Norvig P, Pereira F. The unreasonable effectiveness of data[J]. Intelligent Systems, IEEE, 2009, 24(2): 8-12.

Banko M, Brill E. Scaling to very very large corpora for natural language disambiguation[C]//Proceedings of the 39th Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2001: 26-33.

框架

# 什么是大数据

定义

特点：4v

摩尔定律

Gartner新技术炒作曲线

# 大数据的威力

## 现实生活

塔吉特百货孕妇营销分析，2002年

谷歌预测流感，2009年

奥巴马大选连任成功，2012年

微软大数据成功预测奥斯卡21项大奖，2013年

## 科学研究

机器翻译

语音、图像识别

# 大数据的存储简介

非结构化数据、NoSQL、按列存储、分布式存储

# 大数据的架构简介

## Mapreduce简介

## Spark简介

# 大数据的分析技术简介

协同过滤

主题模型

深度学习

传统机器学习方法并行化：逻辑回归、贝叶斯网络、svm、GBDT、随机森林、knn

# 大数据的应用简介

社交媒体、电子商务、金融、健康医疗、电力、智能交通、电子政务

# 工业大数据简介