



SIGDRM Weekly Letters

SIGDRM 编辑出版

第 171 期

2017-03-06

行业动态

数学家解密如何抵御量子计算的攻击	2
2017 年，大数据发展的二十个新趋势	4
如何为你的机器学习问题选择合适的算法？	11
Twitter 的六大架构设计经验谈	19
15 项新科技成果开启全新互联网世界	27

Hamlin 还表示，“量子计算机在亚原子级的层面上进行操作，理论上讲它的计算能力是普通硅基计算机的数百万倍。所以下一代武装上量子计算机的黑客们在理论上是可以破解当前我们在互联上传输的任何信息的。”

为了打造一个基于未来需要的更好的在线安全系统，Hamlin 和退休数学系教授 William Web 通过采用一种超越普通二进制编码和当前计算机所普遍采用的以 1 - 0 为基准的操作序列的替代数字表达技术，改进了之前的代码，最终在 2015 年写出了“通用背包编码”（Generalized Knapsack Code）。

Hamlin 在他的论文中，用计算机科学家、计算机工程师和其他领域的专家们能共同理解的方式，一步步地详细地阐述“通用背包编码”是如何工作的。

他解释说，通过采用比传统计算机所采用的二进制 0-1 编码技术更为复杂的数串来伪装数据，“通用背包编码”为防御量子计算攻击提供了一套可行的安全方案。

“通用背包编码”通过采用“不仅仅使用 0-1 进行编码的”多种编码方法，扩展了当前计算机所用于操作的二进制编码。

这样的改进，可以阻止更多的网络攻击，这其中也包括可以破解原始版本背包编码的基归约攻击。

Hamlin 希望自己的这篇论文“数学加密技术中的数字”（Number in Mathematical Cryptography）可以帮助扫清人们对于他所研究的、这个处于科学深处的领域的误解，从而使得“通用背包编码”得到更好的发展，并为未来所用。

“量子计算将会改变我们处理数据的方法。但是我们，作为一个社会的整体，我们也必须为这些改变做出相应的准备”，Hamlin 说。

类似于通用“背包编码”这样的编码，在普通常规硬件上的实现，就可以有效地抵御武装上量子计算机黑客的攻击。我想，现在是时候可以考虑在量子计算时代，这些代码在商业和或者通信领域的应用。”

供稿人：向菲

2017 年，大数据发展的二十个新趋势

—来源《百度新闻实验室》

伴随着互联网的深度发展，巨大信息流背后产生的海量数据已成为亟待挖掘的宝藏。2016 即将结束，人们已经在期望新一年的大数据会带来哪些新的变化。下面，我们一起来看看 2017 年大数据发展的二十个新趋势。

“大数据”不再只是一个流行词。弗雷斯特研究公司的研究人员发现，“2016 年，近 40% 的公司在实施大数据技术，并且扩大了采用力度。另有 30% 的公司计划在未来 12 个月内采用大数据技术。”

类似的，NewVantage Partners 的《2016 年大数据高管调查》发现，如今 62.5% 的公司在生产环境中至少有一个大数据项目，只有 5.4% 的企业组织没有计划或开展大数据项目。

研究人员表示，采用大数据技术的势头不太可能很快就减慢。IDC 主管分析和信息管理的集团副总裁丹·维塞特(Dan Vesset)说：“出现的大量数据、新一代技术，以及数据驱动型决策的文化转变，这些因素继续促使市场需要大数据和分

析技术及服务。预计该市场会以 11.7%的复合年增长率继续增长，一直持续到 2020 年。”

虽然大数据市场在增长，但企业组织将如何使用大数据仍不大明朗。新的大数据技术在进入市场，而一些旧技术的使用也在继续增长。

1 数据量将持续增长

数据量的不断增加意味着通过数据的快速分析获取宝贵的市场洞察已经成为大数据业务运营的关键环节。机构和企业组织必须将其内部未被利用的每一字节的大数据，也就是我们所谓的“黑暗数据” (dark data)加以合理化的整合并转化成可以利用的数据资源。

如果大数据还没有为你的企业带来可供战略参考用的新见解，那么在 2017 年记得为你所在的企业提出有关大数据的创新计划，只有这样才能提升企业的竞争优势。

2 利用大数据提升客户体验

对于企业的并购，可以将遗留下来的数据资源转交到分包商系统，这种大数据的使用方式除了可以改进消费者体验之外，还可以升级核心系统。

让消费者使用灵活性的自助服务方式可以让大数据分析为企业快速掌握市场发展的主导趋势，还可以为客户需求增长机遇带来更多有竞争力的市场洞察。

利用大数据更深入的了解客户需求可以让搭配销售或者促销活动提高企业的一线财政收入水平，同时还可以免除因客户流失所导致的业绩缩水风险。

3 Hadoop 的应用领域将更加广泛

将会有越来越多的企业选择采用 Hadoop 和其他类型的大数据存贮架构，相应的，分包商们也将为业主提供更加有创新功能的 Hadoop 解决方案。

当 Hadoop 架构占据有利地位时，企业使用高级分析方法所处理大量数据可以以为盈利决策找到宝贵信息的金矿。

4 预测分析将崭露头角

精准地预测未来可能发生的行为和事件可以提高企业的利润。为降低企业收入风险暴露所使用的欺诈行为快速鉴别和预判技术将会迎来质的飞跃，同时企业运营的卓越性将进一步得到改进。

5 基于云的数据分析将获得更多关注

将数据分析业务迁移到云端可以加速企业采用最新的技术能力，并实现数据资源到行动计划的快速转变。数据分析业务转移到云端之后，企业的运营和技术维护成本也将削减不少。

6 向信息学领域进军并注重数据价值的界定

新的一年，使用信息学助推复杂数据收集、分析与可视化技术的整合可以从数据资源中推导出企业所需的收益来源。从未被充分利用的数据当中提取资源可以提高企业运营绩效。

7 数据可视化将放大商业智能的作用与优势

数据可视化技术让隐藏在大数据资源背后的真相呈现在众人面前。无论数据怎样形成，无论数据资源在哪里，图形数据可视化可以让企业组织在业务繁忙的同时对数据进行检索与处理。

8 物联网、云技术、大数据和网络安全深度融合

数据管理技术，比如说数据质量控制、数据准备、数据分析以及数据整合等方面的融合程度将在新的一年当中达到新的高度。当我们将对智能设备的依赖程度增加时，互通性以及机器学习将会成为保护资产免遭网络安全危害的重要手段。

9 提升数字渠道优化与多渠道体验

以客户偏好的渠道与其保持有效接触可以让企业在传统渠道与数字渠道之间找到最佳平衡点。通过不同渠道不断寻求创新手段提高客户体验度可以带来企业的竞争优势。

10 数据准备和分析的自助式服务将提高效率

无论企业数据类型属于结构化、半结构化还是非结构化，自助服务式的数据预备工具可以加速企业数据准备的时间。使用自助式数据技术可以降低企业对开发团队的依赖程度，从而更重视用户的使用感受，同时企业的运营效率也可以提升。

11 开源

Apache Hadoop 和 Spark 等其他开源应用软件已逐渐主导大数据领域，这个趋势看起来可能会保持下去。一项调查发现，到今年年底，近 60%的企业预计会在生产环境中运行 Hadoop 集群。而据弗雷斯特公司声称，Hadoop 的使用量以每年 32.9%的速度增长。

研究者们表示，2017 年，许多企业会加大使用 Hadoop 和 NoSQL 技术的力度，并想方设法加快大数据处理速度。许多企业会寻求让自己得以实时访问和响应数据的技术。

12 内存技术

许多公司在调查研究，试图加快大数据处理速度的一项技术就是内存技术。

在传统数据库中，数据存储于配备硬盘驱动器或固态硬盘(SSD)的存储系统中。内存技术改而将数据存储于内存中，这大大提高了数据处理速度。弗雷斯特研究公司的一份报告预测，内存数据架构每年会增长 29.2%。

几家厂商提供内存数据库技术，尤其是 SAP、IBM 和 Pivotal。

13 机器学习

随着大数据分析能力不断增强，一些企业已开始投入于机器学习。机器学习是人工智能的一个分支，专注于让计算机可以在没有明确编程的情况下学习新事物。换句话说，它分析现有的大数据存储系统，从而得出可能改变应用程序运行方式的结论。

据 Gartner 声称 机器学习是 2017 年的十大战略技术趋势之一。它特别指出，如今最先进的机器学习和人工智能系统正在超越“基于规则的传统算法，构建能够理解、学习、预测、适应，甚至自主操作的系统。”

14 预测分析

预测分析与机器学习密切相关。实际上，机器学习系统常常为预测分析软件提供引擎。在大数据分析的早期阶段，企业组织回顾数据、查看过去发生了什么，然后开始使用分析工具来调查那些事情为何发生。而预测分析更进了一步，它使用大数据分析工具来预测将来会发生什么。

据普华永道在 2016 年的一项调查显示，如今使用预测分析技术的企业组织数量少得惊人，只有 29%。然而，无数厂商最近推出了预测分析工具，随着公司越来越意识到这种功能强大的工具，这个数字在未来几年可能会急剧提高。

15 智能应用程序

企业使用机器学习和人工智能技术的另一种方式就是构建智能应用程序。这种应用程序常常结合大数据分析技术，分析用户以前的行为，以便提供个性化和更好的服务。大家已经非常熟悉的一个例子就是，现在支持许多电子商务和娱乐应用程序的推荐引擎。

Gartner 在 2017 年的十大战略技术趋势中，将智能应用程序列在第二位。Gartner 副总裁兼研究员大卫·凯洛莱(David Cearley)说：“在今后十年，几乎每个应用、应用程序和服务都会结合某种级别的人工智能。这会成为一种长期趋势，不断演变，并不断扩大人工智能和机器学习在应用程序和服务的应用范围。”

16 智能安全

许多企业还将大数据分析技术纳入到安全战略中。企业组织的安全日志数据提供了以往网络攻击方面的宝贵信息，企业可以利用这些信息来预测、预防和缓解未来的攻击企图。因而，一些企业组织将安全信息和事件管理(SIEM)软件与 Hadoop 等大数据平台整合起来。另一些企业求助于提供的产品整合大数据分析功能的安全厂商。

17 物联网

物联网也可能对大数据产生相当大的影响。据 IDC 在 2016 年 9 月的一份报告声称，“31.4%的受访公司已启动了物联网解决方案，另有 43%期望在今后 12 个月部署这类解决方案。”

随着所有那些新设备和应用程序纷纷上网，企业组织会遇到比过去还要疯狂的数据增长势头。许多企业需要新的技术和系统，以便能够处理和解读来自部署的物联网的潮水般的大数据。

18 边缘计算

边缘计算是一种可以帮助公司处理物联网大数据的新技术。在边缘计算中，大数据分析非常靠近物联网设备和传感器来进行，而不是在数据中心或云端来进行。对企业来说，这带来了一些显著的好处。在网络上传输的数据比较少，这可以改善性能，并节省云计算成本。它让企业组织得以删除只在有限的时间内有价

值的物联网数据 ,从而降低存储和基础设施成本。边缘计算还可以加快分析过程 ,让决策者得以在获得洞察力后比以前更迅速地采取行动。

19 高薪

对 IT 工人来说 ,大数据分析技术的发展可能意味着拥有大数据技能的人才方面需求旺盛 ,薪水优厚。据 IDC 称 : “光在美国 ,2018 年会有 181,000 个深度分析岗位 ,是需要数据管理和解读相关技能的岗位数量的五倍。 ”

由于人才紧缺 , Robert Half Technology 公司预测 ,2017 年数据科学家的平均薪资将提高 6.5% ,年薪在 116000 美元至 163500 美元。同样 ,明年大数据工程师的薪资也将提高 5.8% ,年薪在 135000 美元至 196000 美元。

20 自助服务

由于聘请大数据专家的成本上升 ,许多企业可能寻求让普通专业人员可以满足自己的大数据分析要求的工具。IDC 之前预测 “可视化数据发现工具的增长速度将比商业智能(BI)市场的其余工具快 2.5 倍。到 2018 年 ,投入于支持最终用户自助服务的这种工具将成为所有企业的要求。 ”

几家厂商已经发布了拥有 “自助服务” 功能的大数据分析工具 ,专家预计这个趋势会持续到 2017 年及之后。由于大数据分析变得更加融入到公司所有部门的工作人员的工作方式之中 ,IT 部门可能不太参与到这个过程。

供稿人：杜军毅

如何为你的机器学习问题选择合适的算法？

—来源《THU数据派》

随着机器学习越来越流行，也出现了越来越多能很好地处理任务的算法。但是，你不可能预先知道哪个算法对你的问题是最优的。如果你有足够的时间，你可以尝试所有的算法来找出最优的算法。

本文介绍了如何依靠已有的方法（模型选择和超参数调节）去指导你更好地去选择算法。本文作者为华盛顿大学 eScience Institute 和 Institute for Neuroengineering 的数据科学博士后 Michael Beyeler。

步骤 0：了解基本知识

在我们深入学习之前，我们先重温基础知识。具体来说，我们应该知道机器学习里面三个主要类别：监督学习，无监督学习和强化学习。

在监督学习（supervised learning）中，每个数据点都会获得标注，如类别标签或与数值相关的标签。一个类别标签的例子：将图片分类为「猫」或「狗」；数值标签的例子如：预测一辆二手车的售价。监督学习的目的是通过学习许多有标签的样本，然后对新的数据做出预测。例如，准确识别新照片上的动物（分类）或者预测二手车的售价（回归）。

在无监督性学习（unsupervised learning）中，数据点没有相关的标签。相反，无监督学习算法的目标是以某种方式组织数据，然后找出数据中存在的内在结构。这包括将数据进行聚类，或者找到更简单的方式处理复杂数据，使复杂数据看起来更简单。

在强化学习 (reinforcement learning) 中，算法会针对每个数据点来做出决策 (下一步该做什么)。这种技术在机器人学中很常用。传感器一次从外界读取一个数据点，算法必须决定机器人下一步该做什么。强化学习也适合用于物联网应用。在这里，学习算法将收到奖励信号，表明所做决定的好坏，为了获得最高的奖励，算法必须修改相应的策略。

步骤 1：对问题进行分类

接下来，我们要对问题进行分类，这包含两个过程：

根据输入数据分类：如果我们的数据有标签，这就是一个监督学习问题；如果数据没有标签而且我们想找出数据的内在结构，那这就是无监督学习；如果我们想通过与环境交互来优化目标函数，这是强化学习。

根据输出结果分类：如果模型输出结果是一个数值，这是回归问题；如果输出结果是一个类别，这是分类问题；如果输出结果是一组输入数据，那这是聚类问题。

就是这么简单！

更一般地说，我们可以询问我们自己：我们的算法要实现什么目标，然后以此来找到正确的算法类别。

上面的描述包括了几个我们还没有提到的专业术语：

分类 (classification)：当使用数据来预测类别时，监督学习也被叫做分类。比如将含有「猫」或「狗」的图片识别出来，分类为「猫」或「狗」，这就是二分类问题 (two-class or binomial classification)。当存在更多类别时 (例如预测下一届诺贝尔物理学家的获得者是谁)，这就是所谓的多分类问题 (multi-class classification)。

回归 (regression) : 当要预测数值时 (比如预测股价) , 监督学习也被称为回归。

聚类 (clustering) : 聚类或聚类分析 (cluster analysis) 是无监督学习中最常见的方法之一。聚类是将一组对象以某种方式分组, 使得同一组中的数据比不同组的数据有更多的相似性。

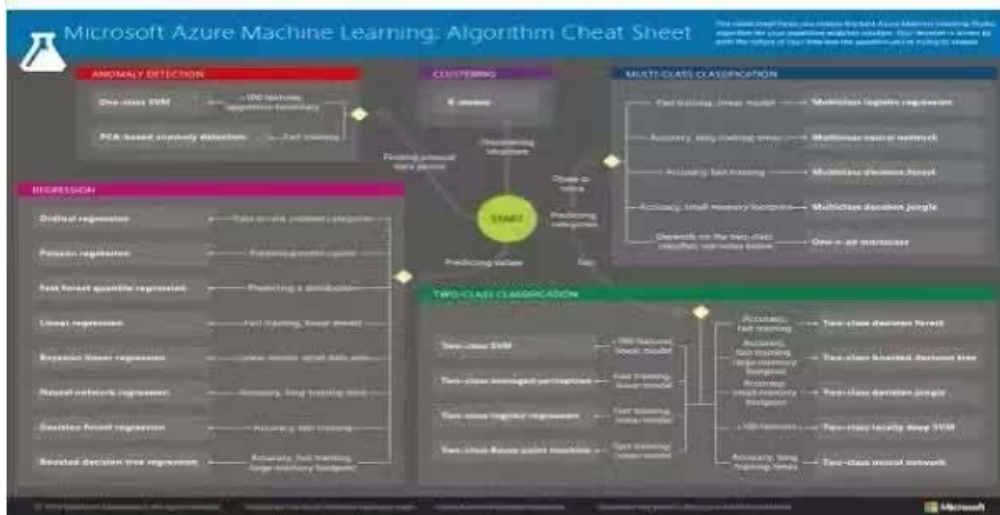
异常检测 (Anomaly detection) : 有时我们需要找出数据点中的异常点。

例如, 在欺诈检测中, 任何极不寻常的信用卡消费都是可疑的; 欺诈具有大量不同的形式, 而训练样本又非常少, 使得我们不可能完全了解欺诈活动应该是什么样。异常检测所采取的方法就是了解正常情况下的表现行为 (使用非欺诈交易的历史数据) , 并识别出显著不同的表现行为。

步骤 2 : 寻找可用的算法

现在我们已经将问题进行了分类, 我们就可以使用我们所掌握的工具来识别出适当且实用的算法。

Microsoft Azure 创建了一个方便的算法列表, 其展示了哪些算法可用于哪种类别的问题。虽然该表单是针对 Azure 软件定制的, 但它具有普遍的适用性 (该表单的 PDF 版本可查阅 <http://suo.im/3Ss2zW>) :



一些值得注意的算法如下：

分类：

支持向量机（SVM）可用于找到尽可能宽的分类的边界。当两个分类不能被清楚地分开时，该算法会找到其所能找到的最佳边界。其真正的亮点在于处理特征密集的数据，比如文本或者基因组（特征数量 > 100）。

在这些情况下，除了仅需要适量的记忆外，支持向量机（SVM）能够比其它大多数算法更快且更少过拟合地进行分类。

人工神经网络是涵盖二分类、多分类和回归问题的脑启发式学习算法。

它们有无限种类，包括感知器和深度学习。它们需要很长时间来训练，但已知其在多种应用领域都实现了当前最佳的表现。

logistic 回归：即便名字中有着「回归」，但 logistic 回归实际上是一种可用于二分类和多分类问题的强大工具。

它快速且简单。事实上，它使用「S」形曲线而非直线，所以它自然适合于数据分组。logistic 回归可以给出线性分类边界，所以如果你要使用它，你一定要确保你能接受线性的近似。

决策树和随机森林：决策森林 (decision forests) (回归、二分类、多分类) , 决策丛林 (decision jungles) (二分类和多分类) 和提升决策树 (boosted decision trees) (回归和二分类) 都基于决策树。

这是一个基本的机器学习概念。决策树有许多不同的变体，但它们都在做同样的事情—将特征空间 (feature space) 细分为具有大致相同标签的区域。这些区域可以是一致的类别或者恒定值，具体取决于你进行的是分类还是回归。

回归：

线性回归是将一条线 (或平面、或超平面) 拟合到一个数据集上。这是一种主要的工具，简单且快速，但对于一些问题而言，它可能过于简单。

贝叶斯线性回归有着非常理想的特性：它可以避免过拟合。贝叶斯方法通过事先对答案的可能分布做出一些假设来做到这一点。这种方法的另一个副产品是它们具有非常少的参数。

提升决策树回归 (Boosted decision tree regression)：如上所述，提升决策树 (回归和二分类) 均基于决策树，并通过将特征空间细分为具有大致相同标签的区域发挥效用。

提升决策树通过限制其可以细分的次数以及每个区域中所允许的最少数据点来避免过拟合。该算法会构造一个树的序列，其中每棵树都会学习弥补之前的树留下来的误差。这能得到一个会使用大量的内存的非常精确的学习器。

聚类：

层次聚类 (Hierarchical Clustering) 的目标是构建聚类的层次结构，它有两种形式。聚集聚类 (agglomerative clustering) 是一种「自下而上」的方法，其中

每个观察 (observation) 在其自己的聚类中开始，随着其在层次中向上移动，成对的聚类会进行融合。

分裂聚类 (divisive clustering) 则是一种「自上而下」的方法，其中所有的观察都从一个聚类开始，并且会随观察向下的层次移动而递归式地分裂。整体而言，这里的融合和分裂是以一种激进的方式确定的。层次聚类的结果通常表示成树状图 (dendrogram) 的形式。

k-均值聚类 (k-means clustering) 的目标是将 n 组观测值分为 k 个聚类，其中每个观测值都属于其接近的那个均值的聚类——这些均值被用作这些聚类的原型。这会将数据空间分割成 Voronoi 单元。

异常检测：

k 最近邻 (k-nearest neighbors / k-NN) 是用于分类和回归的非参数方法。在这两种情况下，输入都是由特征空间中与 k 最接近的训练样本组成的。

在 k-NN 分类中，输出是一个类成员。对象通过其 k 最近邻的多数投票来分类，其中对象被分配给 k 最近邻中最常见的类 (k 为一正整数，通常较小)。在 k-NN 回归中，输出为对象的属性值。该值为其 k 最近邻值的平均值。

单类支持向量机 (One-class SVM)：使用了非线性支持向量机的一个巧妙的扩展，单类支持向量机可以描绘一个严格概述整个数据集的边界。远在边界之外的任何新数据点都是非正常的，值得注意。

步骤 3：实现所有适用的算法

对于任何给定的问题，通常有多种候选算法可以完成这项工作。那么我们如何知道选择哪一个呢？通常，这个问题的答案并不简单，所以我们必须反复试验。

原型开发最好分两步完成。在第一步中，我们希望通过最小量的特征工程快速且粗糙地实现一些算法。

在这个阶段，我们主要的目标是大概了解哪个算法表现得更好。这个步骤有点像招聘：我们会尽可能地寻找可以缩短我们候选算法列表的理由。

一旦我们将列表减少至几个候选算法，真正的原型开发开始了。理想情况下，我们会建立一个机器学习流程，使用一组经过仔细选择的评估标准来比较每个算法在数据集上的表现。

在这个阶段，我们只处理一小部分的算法，所以我们可以把注意力转到真正神奇的地方：特征工程。

步骤 4：特征工程

或许比选择算法更重要的是正确选择表示数据的特征。从上面的列表中选择合适的算法是相对简单直接的，然而特征工程却更像是一门艺术。

主要问题在于我们试图分类的数据在特征空间的描述极少。例如，用像素的灰度值来预测图片通常是不佳的选择；相反，我们需要找到能提高信噪比的数据变换。

如果没有这些数据转换，我们的任务可能无法解决。例如，在方向梯度直方图（HOG）出现之前，复杂的视觉任务（像行人检测或面部检测）都是很难做到的。

虽然大多数特征的有效性需要靠实验来评估，但是了解常见的选取数据特征的方法是很有帮助的。这里有几个较好的方法：

主成分分析（PCA）：一种线性降维方法，可以找出包含信息量较高的特征主成分，可以解释数据中的大多数方差。

尺度不变特征变换 (SIFT) : 计算机视觉领域中的一种有专利的算法, 用以检测和描述图片的局部特征。它有一个开源的替代方法 ORB(Oriented FAST and rotated BRIEF) 。

加速稳健特征 (SURF) : SIFT 的更稳健版本, 有专利。

方向梯度直方图 (HOG) : 一种特征描述方法, 在计算机视觉中用于计数一张图像中局部部分的梯度方向的 occurrence。

更多算法请参考 : https://en.wikipedia.org/wiki/Visual_descriptor

当然, 你也可以想出你自己的特征描述方法。如果你有几个候选方法, 你可以使用封装好的方法进行智能的特征选择。

前向搜索 :

最开始不选取任何特征。

然后选择最相关的特征, 将这个特征加入到已有特征 ; 计算模型的交叉验证误差, 重复选取其它所有候选特征 ; 最后, 选取能使你交叉验证误差最小特征, 并放入已选择的特征之中。

重复, 直到达到期望数量的特征为止 !

反向搜索 :

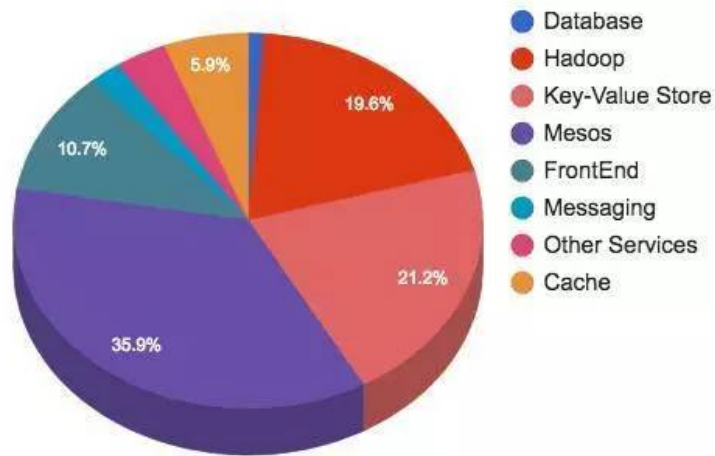
从所有特征开始。

先移除最不相关的特征, 然后计算模型的交叉验证误差 ; 对其它所有候选特征, 重复这一过程 ; 最后, 移除使交叉验证误差最大的候选特征。

重复, 直到达到期望数量的特征为止 !

使用交叉验证的准则来移除和增加特征 !

步骤 5 : 超参数优化



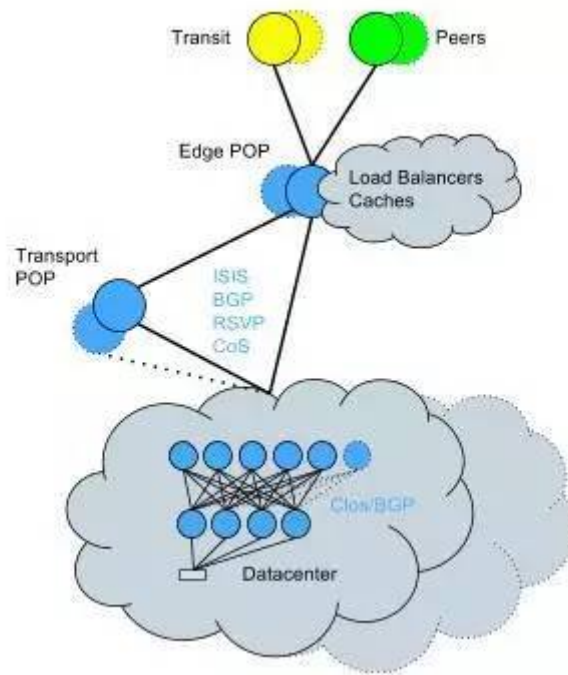
网络流量

2010 年初，Twitter 团队开始考虑将集群从第三方主机上迁出，这个决定和动作意味着 Twitter 团队需要学习如何构建和运行内部的基础设施，由于需要在有限的可视化情况下了解核心基础设施需求，团队开始调研各种网络设计、硬件，以及供应商。

到 2010 年下半年，Twitter 团队完成了第一个网络体系结构设计，解决了科罗拉多主机集群遇到的扩展性和服务问题。该方案有深度缓冲设计，支持对于突发的流量请求以及确保电信核心交换机在网络层没有超载。这个方案支撑了 Twitter 的早期版本，创造了一些比较出名的业绩，例如打破了 TPS 数据记录的“天空之城”事件（每秒处理 34000 条记录）以及应对 2014 年世界杯。

此后的几年时间里，Twitter 的数据中心运行在五大洲的成千上万的服务器，网络覆盖很广。从 2015 年上半年开始，Twitter 开始遭受到了成长过程中的痛苦，由于不断变化服务系统架构和增加容量需求，最终达到了数据中心物理可扩展性的上线，网状拓扑结构不再支持通过增加新的机架提升性能，即不再支持增加额外的硬件。另外，Twitter 现有的数据中心开始由于不断增加路由规模和复杂的网络拓扑结构导致出现不稳定异常情况。

为了解决这个问题，Twitter 开始将现有的数据中心转换为 Clos 拓扑+BGP，这是一种必须在现场做的网络转换工作。尽管很复杂，但是 Twitter 在一个相对较短的时间内完成了这个转换，并且对服务影响最小。网络拓扑图看起来是这样的：



总结技术亮点如下：

单个设备故障具有较小的影响范围。

水平带宽缩放能力。

路由引擎较低的 CPU 开销；路由更新更高效的处理能力。

由于较低的 CPU 开销，更高的路由容量。

每个设备和链路上的更细粒度的路由策略控制。

不再重复发生之前的几个已知问题：包括增加协议收敛时间、路线流失问题和伴随固有的 OSPF 复杂性的不可预期问题。

启用无影响机架迁移方案。

数据中心

Twitter 的第一个数据中心是基于建模分析出的能力和已知系统的运行状态经验构建的。但是仅仅几年之后，数据中心比最初的设计扩大了 400%。现在，随着 Twitter 应用程序堆栈的演变，Twitter 正在变得更加分布式化，运行状态也在跟着变化。引导 Twitter 进行网络设计的最初假设场景已经不复存在了。

业务需求增长过快，导致针对整个数据中心进行重构已经不切实际。所以构建一个高可扩展的体系架构会让 Twitter 更容易增加能力，而不是采用叉车式的迁移方案。

高可扩展性的微服务需要高可靠性网络，可以支持处理各种业务。Twitter 的业务范围从 TCP 长连接到离线的 MapReduce 任务，再到超短连接。针对这类多样性业务需求的应对方案是，部署具有深包缓冲区的网络设备，但是这样会带来一系列问题：更高的成本和更好的硬件复杂度。之后的设计 Twitter 使用了更加标准化的缓冲区大小，以及在提供切断开关功能的同时，提供了更好的 TCP 栈服务器，这样可以更好的处理网络风暴问题。

骨干网络

Twitter 的骨干网流量每年都有大幅度正常，并且仍然可以看到数据中心之间的突发数据增长较正常状态的 3-4 倍情况存在。这种情况对于老的协议是一个特有的挑战，这些老的协议，例如 MPLSRsvp 协议，从来不是为应对突然爆发的网络风暴而设计的，它的目标是应对渐进式的缓慢的网络请求增长。为了获得尽可能快的响应时间，Twitter 不得不花费大量的时间调整这些协议。此外，Twitter 实现的优先次序可以处理网络高峰（特别是存储复制）情况。

Twitter 需要确保在任何时候优先客户的传输流量，可以通过延迟低优先级的存储复制工作满足这个需求，存储复制这类工作有一天时间的 SLA。这样就可以使用最大量的网络资源，让数据尽可能快速地移动。客户业务需求比低优先级的后台业务需求优先级高。而且，为了解决伴随着 RSVP 自动带宽而来的 bin-packing 问题，Twitter 实现了 TE++，这个工具当流量增加时创建额外的 LSP，当流量下降时则会删除 LSP。这使得 Twitter 可以有效地管理连接之间的网络业务，同时减少维护大量的 LSP 所造成的 CPU 负担。

然而主干网从最初开始就缺少任何业务工程设计，后来增加了帮助 Twitter 可以根据业务增长进行扩展的特性。为了实现这一点，Twitter 完成了角色分离，使用单独的路由器分别处理核心和边缘路由请求。这也使得 Twitter 能够低成本效益地扩展，而不需要购买复杂的带边缘功能的路由器。

在边缘路由层，Twitter 有一个核心连接所有网络，并且可以支持水平扩展，例如每个站点有很多路由器，不止两个，因为 Twitter 有一个核心层互联所有的设备。

为什么扩展路由器的 RIB(路由信息库，即 Routing InformationBase)，Twitter 引入了路由反射机制，这样就可以适配扩展需求，但是在做这个的过程中，需要移植到变更设计，Twitter 也做了路由反射的客户端！

存储设计

每天有数以百万计的的推文（微博）被发送出来。这些推文需要被处理、存储、缓存、服务以及分析。对于如此庞大的信息内容，Twitter 需要一个稳定的基础设施。存储和消息占据了 45% 的 Twitter 的基础设施空间。

存储和消息团队提供了如下服务：

用于运行计算和 HDFS 的 Hadoop 集群

用于所有的低延迟键值存储的 Manhattan 集群

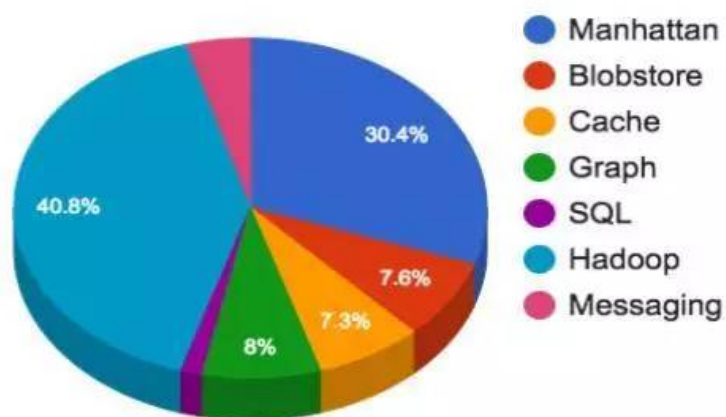
Graph 用于存储 MySQL 集群存储分片

Blobstore 集群用于所有的大型对象（视频、图片、二进制文件等等）

缓存集群

消息集群

关系型存储（MySQL、PostgreSQL 以及 Vertica）



在这个规模的多租户领域，Twitter 遇到了一些困难，其中有一个是必须解决的问题。很多时候客户有一些很奇怪的用例，这些用例实施后会影响到其他租户，这种案例带来了专有集群设计。许多专有集群增加了运行的工作量用于保持程序运行。

Twitter 的集群设计没有什么特别的，不过倒是有一些有意思的地方可以分享：

Hadoop：Twitter 有多个集群存储了超过 500PB 的数据，这些数据被分为四组（实时数据、处理数据、数据仓库，以及冷数据仓库）。Twitter 最大的集群有超过 1 万个节点组成。该集群每天运行 15 万个应用程序，启动 13 亿个容器。

Manhattan (Tweets、直接消息、Twitter 账号以及其他一些东西的后端) :
Twitter 运行了一些集群针对不同的用例，例如大型多租户、小型的非常用功能、只读的，以及针对重写/重读业务模式下的读写。只读集群可以处理几千万 QPS，而读/写集群可以处理百万级的 QPS。Twitter 观察到的性能最好的集群每天处理几十万的写请求。

Graph : Gizzard/MySQL 这样的基于分片的集群用于存储 Twitter 的图片。这个集群可以管理顶峰时间段千万级别的 QPS，平均到每台 MySQL 服务器大约 3 万-4.5 万 QPS。

Blobstore : Twitter 的图片、视频，以及大型文件的存储量已经达到了数以百亿计。

Cache : Redis 和 Memcache 集群，缓存 Twitter 用户、时间表、推文，以及其他一些内容。

SQL : 包括 MySQL、PostgreSQL 和 Vertica。MySQL/PostgreSQL 被用于强一致性场景，就像内部工具一样管理广告活动、广告切换。

Hadoop/HDFS 也是日志管道的后端存储，但是在过渡到 Apache Flume 的最后的测试阶段，作为一种替换方案解决了选择客户端聚合时缺乏速率限制、缺乏传递类型保证等局限性，并且解决了内存奔溃问题。Twitter 每天处理一兆条信息，并且所有的信息被处理后超过 500 个类别，然后有选择地在所有集群内拷贝。

缓存设计

虽然缓存只占用基础设施的 3%，但是它确实对于 Twitter 很关键。缓存保护后端存储层远离业务风暴，允许存储流动成本很高的对象。Twitter 在巨型规模内使用了几款缓存技术，例如 Redis 和 Twemcache。更具体地说，Twitter 有一个

专用和多租户混用的 Memcached (twemcache) 集群，以及夜鹰集群 (分布式 Redis)。Twitter 也迁移了几乎所有的主要的缓存到 Mesos，以降低运行成本。

对于 Cache 来说，扩展性和性能是首要挑战。Twitter 运营了数个集群，总计每秒 320M 数据包的数据包速率，提供超过 120GB/s 的数据给客户，Twitter 的目标是即使在高峰段，也要确保每次响应的延迟约束在 0.1% 到 0.01% 之间。

为了满足 Twitter 的高吞吐量和低延迟服务水平目标 (SLO)，Twitter 需要持续测试系统性能，寻找更有效的优化方案。为了帮助内部做到这一点，Twitter 写了 rpc-perf 工具，更好地理解缓存系统是如何工作的。当 Twitter 尝试从专有机器上迁移到现在的 Mesos 基础设施时这个工具的作用就很重要了，帮助制定了容量规划。这些优化努力的结果是 Twitter 在没有延迟的基础上每台机器的吞吐量提升了一倍以上。Twitter 仍然坚信存在很大的调优空间。

Twitter 作为一家互联网服务提供商，它在建设通信软件服务时遇到的网络问题、软件系统架构问题、软件技术选型等等都值得我们学习。Twitter 发表的文章有助于读者从整体理解互联网软件开发、发布、问题解决等整个体系结构相关知识，也帮助读者从侧面完成技术选型。

供稿人：王小雪

镜”“Transistor Density Increase by 1000X”“三星复合生物信号处理器”“SAP 工业 4.0 互联制造解决方案”“华为麒麟 960 手机 soc 芯片”和“Qualcomm 5G NR 原型系统和试验平台”。这些都属互联网领域“最新的理论研究、最前沿的技术成果、最具影响力的产品和最具创新性的商业模式”。

中国工程院院士、中国互联网协会理事长邬贺铨认为，这些领先的科技成果不仅是一种技术或者平台，更表现出一种国际的思想、领导力。

微软的“HoloLens 全息眼镜”探索的是“虚拟的数字世界和物理的现实世界的无缝融合”。微软全球执行副总裁沈向洋说，HoloLens 全息眼镜是一个惊人的创新，是世界上最先进的全息计算机。

HoloLens 全息眼镜搭载了 Windows10 操作系统，完全独立运营。“HoloLens 全息眼镜拥有难以置信的计算能力，众多的传感器实时扫描进行三维建模，还有高清相机。”沈向洋说，通过 HoloLens 来探索火星，利用火星探测器的全息影像，科学家可以身临其境般在火星表面工作。在机械制造和设计方面，HoloLens 不仅可以呈现发动机的三维全息模型，还可以在其之上进行零部件的叠加，甚至是透视其内部结构。

而由超大规模神经网络、计算以及数据三大部分组成的“百度大脑”，包含了语音、图像、自然语言处理、用户画像等能力。其中，百度深度语音识别技术 Deep Speech 2 入选《麻省理工科技评论》“2016 年十大突破技术”，语音识别准确率达到 97%；百度多项人脸检测识别、OCR 竞赛均世界第一；情感分析、句法分析、命名实体识别等技术也在国际标准评测集上获世界第一。

百度公司总裁张亚勤介绍，“百度大脑”以其领先算法、超强计算能力、百亿级大数据，而被广泛应用于无人车、智能医疗等方面。

世界互联网领先科技成果评选由世界互联网大会组委会发起，旨在展现全球互联网领域最新科技成果，扩大互联网创新力量的影响力，彰显互联网领域从业者的创造性贡献，搭建全方位的创新交流平台。

为此，大会组委会组建了包括科技部部长万钢，中国工程院院士、中国互联网协会理事长邬贺铨，前美国总统科技顾问、2004年图灵奖获得者罗伯特·卡恩等在内的33位海内外知名的互联网专家组成的推荐委员会。

经过两个月紧张评选，最终推荐委员会从来自亚洲、非洲、欧洲、北美洲、南美洲、大洋洲的500多项互联网领域的领先科技成果中，推荐遴选出最具代表性的15项领先科技成果。

中科院计算所所长孙凝晖表示，如果明年世界互联网大会继续评选互联网领域的领先科技的话，希望可以看到有关云计算、大数据和互联网安全的技术出现。

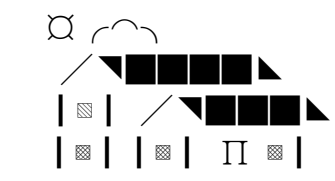
在他看来，包括人工智能在内的互联网技术的共同突破，一定会开启未来全新的互联网世界，一个更懂用户的互联网世界。

本报乌镇11月16日电

供稿人：韩林茜

» SIGDRM WEEKLY LETTERS «

» QQ: 10006549 «



主 编：张志勇
副 主 编：赵长伟
责任编辑：史培宁
网络校对：韩林茜

• D R M° •