

在大数据时代的车牌汉字识别方法

● 沈文枫 张建蕾 周丁倩 陈圣波 邱峰

上海大学 计算机工程与科学系 上海 200444

摘要：

在大数据时代，交通成为网络数据量最大的数据来源之一，智能交通成为必然需求。车牌识别是智能交通的基础，可被广泛的应用在了车库，交通监控等工程上，然而识别的准确率却有待加强，已有的算法对于字母、数字的识别准确率都非常高，而对于中国特有的汉字识别却效果不佳。提出用受限玻尔兹曼机组成的深信度网络算法来识别车牌字符，大大提升了汉字识别的准确率，准确率达到99.44%。

关键字：车牌汉字识别，深信度网络，限制玻尔兹曼机，深度学习

在大数据时代，交通成为网络数据量最大的数据来源之一，每天每个地市的每个十字路口要过数以万计的车，各种车型，各种牌照。而且随着科技和经济的高速发展，购置汽车的人与日俱增，然而日益增多的车辆在给人们的出行带来便捷的同时也这造成了相应的麻烦，交通问题日益突出，这使得发展智能交通系统（Intelligent Transportation System，简称ITS）成为需要迫切解决的问题。如交通堵塞和交通事故等问题。借助智能交通系统^[1]能够实现交通的智能化管理，进而实现对交通环境的全面有效的监管，帮助缓解交通堵塞、交通安全等方面出现的问题。所以智能交通系统的研发迫在眉睫。

车辆牌照识别（License Plate Recognition简称LPR）技术^[2]作为ITS的基础，广泛地应用于违章车辆稽查、城市监控系统、停车场、高速公路收费站等需要认证车牌的相关场合。其运用图像分割和识别理论，分析处理汽车牌照号码图案以确定其位置，并通过进一步提取和识别获取牌照字符信息，为科学有效的识别和管理车辆发挥了积极的作用，其重要性不言而喻。

车牌字符^[3]包含汉字、字母和数字，比起字母和数字，汉字笔画多并且结构复杂，因此，汉字识别是车牌字符识别的难点和重点。近几年，国内研究者提出了多种车牌字符识别方法，这些方法对字母和数字的识别率高于汉字。因此研究车牌中汉字

识别方法，提高车牌汉字的识别率具有理论意义和应用价值。

从车牌中提取的汉字易受车牌提取过程中环境以及车牌本身完整性及清洁程度的影响，因此会存在模糊、污损及变形等问题，增加了识别难度。本文针对这些问题，对车牌中的汉字识别方法展开探讨，主要贡献如下。

（1）车牌汉字预处理：由于每次从车牌中切割出来的汉字图像中，汉字所占的大小和所处的位置均不相同，因此，需要对图像进行大小归一化和位置归一化处理。将汉字图像统一调整为（28维*28维）大小，并保证汉字的大小和位置相同。

（2）汉字特征一维化：按汉字的大小（28维*28维）可将汉字一维化为一个784维特征向量，也是本研究所使用的深度学习网络的输入层神经元个数。

（3）深度学习识别器设计：由于深度学习具有较强的抗干扰能力、容错能力和自学习能力，尝试利用深信度网络进行用汉字的特征向量进行训练，而后是用训练好的模型进行汉字识别。与神经网络方法相比，采用深度学习方法具有更高的识别率和抗干扰能力。

1. 相关工作

自动车牌识别是交通领域的一大应用。虽然车

牌定位技术逐渐成熟, 车牌上的字符识别准确率却远未满足。车牌的字符识别难处之一就是字符的特征提取, 而国外的车牌都是由字母与数字组成, 笔画简单, 经过图像处理后的形变量小, 易识别。Wu等^[4]用BP神经网络算法识别车牌, 字母识别准确率达到99.25%, 数字的识别准确率达到99.3%。Cheng等^[5]利用一种混合算法对中国车牌中的汉字进行识别, 准确率达到95.96%。中国的车牌识别准确率关键在于对字符的识别。而在车牌字符识别中, 与字母、数字识别相比, 汉字的识别率是最低的, 所以提高汉字的识别准确率, 对于车牌识别的准确率提高将有一个很大的提升。由于汉字的笔画多, 笔画间的空隙小, 经过处理后会出现少笔画、多笔画、断笔、一块黑、一块白的情况。如图1所示: (字分别为: a.赣; b.津; c.京; d.鲁; e.鲁; f.闽; g.苏; h.皖; i.豫; j.粤), 这使得车牌汉字识别的难度增大。

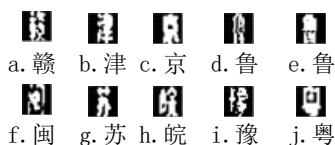
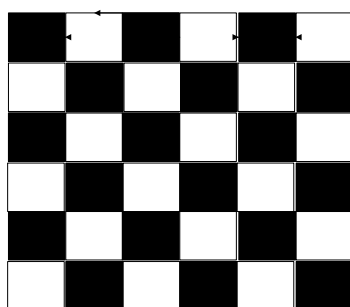
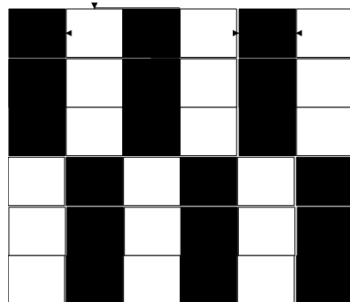


图1 摄像头拍摄的车牌汉字样本

最常用的直方图特征算法^[6]提取特征时, 易出现两个纹理完全不同的图提取后其特征值一样的问题(见图2)。



A 水平特征值 $h=(3, 3, 3, 3, 3)$
垂直特征值 $v=(3, 3, 3, 3, 3)$



B 水平特征值 $h=(3, 3, 3, 3, 3)$
垂直特征值 $v=(3, 3, 3, 3, 3)$

图2 直方图特征算法

而汉字的特征比较复杂, 很多字的直方图特征

差异不大, 而有的相同字的样本的直方图差异却特别大, 如图1中的“d.鲁”和“e.鲁”。直方图特征提取方法在识别数字和字母时, 由于它们的笔画少、纹理容易辨别, 所以准确率很高; 但在汉字识别上, 由于汉字的结构复杂, 笔画多且密集, 在将从拍摄的图片处理成汉字二值图时, 已经成缺笔画或黑白不均的现象, 使得不同汉字的直方图算法特征值可能相同, 相同汉字的直方图算法特征值却可能不同, 说明该算法的鲁棒性较差。

在很多情况下, SIFT(Scale-invariant feature transform)特征提取方法^[7]被用来定位车牌, 但该算法在车牌的特征比较明显简单(矩形, 有七个字符等特征)的情况下, 识别效果比较理想。而对于汉字来说, 由于拍摄的车牌汉字经处理后与正规车牌汉字差异较大, 所以它的识别效果不理想。

神经网络分类算法^[4,8]的分类效果较好, 但依赖于有效的特征提取方法, 如果直接用图片像素特征直接进行分类, 则训练参数会非常大, 训练时间与训练难度会大大增加, 且容易出现分类错误或者过拟合情况。另一方面, 由于车牌拍摄的角度, 字符分割等情况, 每个字符图像中字符的位置和倾斜程度不一, 所以同一个汉字的不同样本其像素值分布差异大, 造成神经网络算法对于拍摄车牌的汉字的识别准确率难以上升。所以对于神经网络方法不太适合用来做车牌的汉字识别。

Google的开源软件Tesseract^[9], 是OCR(Optical Character Recognition)领域里看的一款识别软件, 对于打印的图片中字符(包括汉字)的识别率可以达到99.99%以上, 但是对于手写的字或者摄像头拍摄的车牌的汉字识别, 其测试准确率非常低, 甚至不到80%, 所以对于非规则字体, Tesseract的识别准确率大幅下降。

综上所述, 各种算法都存在一定的缺陷, 使得对于中国车牌特有的汉字的识别率不是很高。本文中, 采用的方法是深度学习中的深信度网络算法, 它由3个玻尔兹曼机和1个逻辑回归网络构成。

2. 车牌汉字识别网络设计

深度学习^[10,11]是模仿人脑进行分析学习的神经网络, 试图模拟人脑的行为特征, 对数据进行分析与解释。深度学习网络通常含有多个隐藏层, 其排列分布与人脑相类似。它通过对较低层特征的组合、分析、筛选, 形成较高层的特征, 以发现数据的分布式特征表示。

深信度网络^[12,13]是深度学习的一种, 通常由数个特征提取层(RBM)以及一个全连接的分类器组成。它的权值共享网络结构使之更类似于生物神经

网络。其具有更低的模型复杂度与更少的权值，从而可以直接输入目标图像，省去了繁琐的特征提取过程；将图像的局部作为最低层的输入，从而获取到输入图像最为基础的特征，所以对目标物体或图像的形变、旋转具有较好的抵抗性。与传统的BP神经网络相比，它具有训练参数少、可直接输入二维图像等优点。受限玻尔兹曼机是深信度网络的主要构成部分，通过逐层特征提取，最后得到想要的结果。

2.1 受限玻尔兹曼机

受限玻尔兹曼机^[14,15]是由Hinton和Sejnowski于1986年提出的一种生成式随机神经网络(generative stochastic neural network)，该网络由可见层和隐藏层构成，各层的变量都是二值变量，亦即其状态取{0,1}。整个网络是一个二部图，只有可见单元和隐藏单元之间才会存在边，可见单元之间以及隐藏单元之间都不会有边连接，如图3所示：

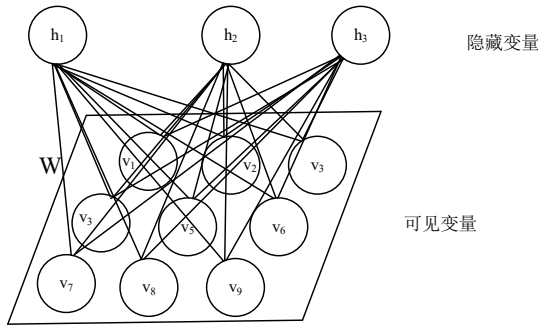


图3 受限玻尔兹曼机原型

图3所示的RBM含有9个可见单元(构成一个向量 $V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9\}$)和3个隐藏单元(构成一个向量 $H = \{h_1, h_2, h_3\}$)，每个边都有权值，共 9×3 个边，用 W 矩阵表示，也就是可见单元和隐藏单元之间的边的权重。

RBM是一种基于能量(Energy-based)的模型，其可见变量 v 和隐藏变量 h 的联合配置(joint configuration)的能量为

$$E(v, h; \theta) = - \sum_{i=1}^D \sum_{j=1}^F W_{ij} v_i h_j + \sum_{i=1}^D v_i b_i + \sum_{j=1}^F h_j a_j \quad (1)$$

其中 θ 是RBM的参数 $\{W, a, b\}$ ， W 为可见单元和隐藏单元之间的边的权重， b 和 a 分别为可见单元和隐藏单元的偏置(bias)。

有了 v 和 h 的联合配置的能量之后，就可以得到 v 和 h 的联合概率为

$$P_{\theta}(v, h) = \frac{1}{Z(\theta)} \exp(-E(v, h; \theta)) \quad (2)$$

其中 $Z(\theta)$ 是归一化因子，其函数原型为 $Z(\theta) = \sum_{v,h} \exp[-E(v, h; \theta)]$ 。

通过最大化观测数据的似然函数 $P(v)$ ，可以得到RBM收敛， $P(v)$ 可由式(2)求 $P(v, h)$ 对 h 的边缘分布得到：

$$P_{\theta}(v) = \frac{1}{Z(\theta)} \sum_h \exp[v^T W h + a^T h + b^T v] \quad (3)$$

最大化 $P(v)$ 后就可以得到RBM的参数。最大化 $P(v)$ 等同于最大化 $\log P(v) = L(\theta)$ ：

$$L(\theta) = \frac{1}{N} \sum_{n=1}^N \log P_{\theta}(v^{(n)}) \quad (4)$$

N 为RBM个数。这样，通过训练求得RBM的参数就可得到隐藏层，即特征提取层。

B、深信度网络

深信度网络^[15]是一个概率生成模型，与传统的判别模型的神经网络不同，生成模型是建立一个观察数据和标签之间的联合分布，对已知标签求观察数据的概率 $P(L|O)$ (Label|Observation) 和已知观察数据求标签的概率 $P(O|L)$ 都做了评估，而深信度网络的判别模型仅仅评估了后者，也就是 $P(O|L)$ 。

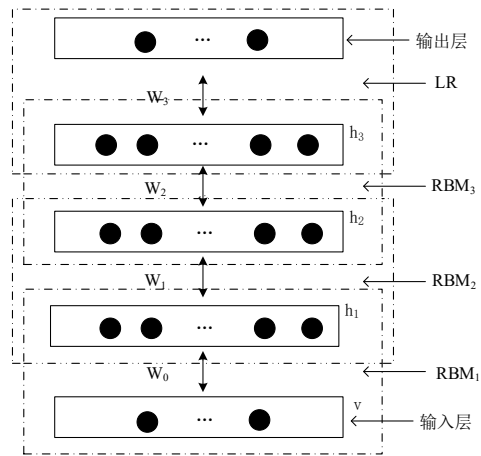


图4 深信度网络构成部件

能量公式 $P(v, h_1, h_2, h_3) = P(v|h_1)P(h_1|h_2)P(h_2|h_3)$

它是由三个限制玻尔兹曼机 (Restricted Boltzmann Machines) 层组成，这些网络被“限制”为一个可视层和一个隐层，层间存在连接，但层内的单元间不存在连接。隐层单元被训练去捕捉在可视层表现出来的高阶数据的相关性。

最开始的时候，通过一个非监督贪婪逐层方法去预训练获得生成模型的权值，非监督贪婪逐层方法被Hinton等^[13]证明是有效的，并被其称为对比分歧 (contrastive divergence)。在这个训练阶段，在可视层会产生一个向量 V ，通过 V 将可视层神经元的值传递到隐藏层。反之，可视层的输入会被随机选择，以尝试重构原始输入信号。新的可视的神经元激活单元将前向传递重构隐藏层激活单元，获得向量 H 。这些后退和前进的步骤就是Gibbs采样^[16]，而隐藏层激活单元和可视层输入之间的相关性差别就作为权

值更新的主要依据。

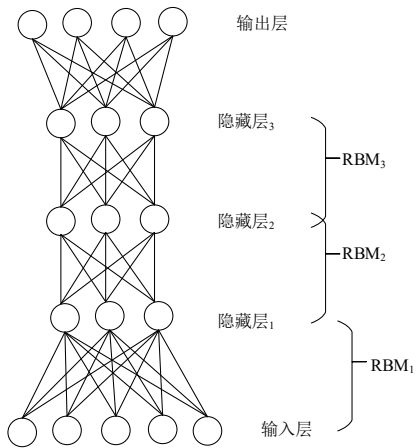


图5 深信度网络架构

经预训练后，DBN可利用带标签数据通过逻辑回归算法去对判别性能做调整。这里，一个标签集将被附加到顶层，通过自下向上的学习到的识别权值获得一个网络分类面。该性能比单纯的逻辑回归算法训练的网络更好。这可以很直观的解释为DBNs的逻辑回归算法只需要对权值参数空间进行一个局部的搜索，这比前向神经网络的训练时间更快，而且收敛的时间也较少。

3. 车牌汉字识别实验

首先，实验数据是来自摄像头拍摄的车牌，通过采集大量车牌，然后从中抠出汉字，形成训练集与测试集，数据来源过程如图6所示。



图6 拍摄车牌图片的汉字提取过程

将从大街上拍到的车牌就经过上述处理后，每个车牌可得一个汉字样本。共收集到5000张车牌图片，其中样本数比较少的车牌汉字样本（“川”，“赣”等）也有十多个，样本数比较多的车牌汉字样本（“京”，“沪”）有几百个，已经够成可训练和可测试的数据集。在中国，车牌汉字样本种类“军北南广沈成兰济空海京津冀晋蒙辽吉黑沪苏浙皖闽赣鲁豫鄂湘粤桂玉琼川贵云藏陕甘青宁新”共41个汉字。个别的“空”、“海”等汉字由于数量太少而不做训练，其它的共34个字经训练处理后进行测试，验证识别准确率。

将所得的样本重置大小为（28维*28维），再

进行一维化后生成一个784维向量，构成输入层，共有784个神经元。此时，输入层和隐藏层1构成一个RBM1，输入层是RBM1的可视层，隐藏层1为RBM1的隐藏层。同理，RBM1的隐藏层为RBM2的可视层，隐藏层2为RBM2的隐藏层，如下图所示。

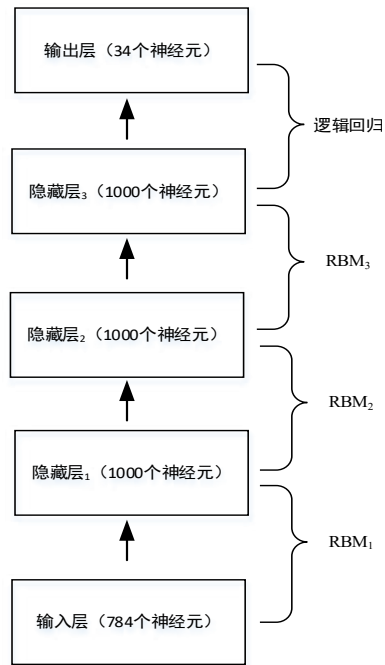


图7 深信度网络实验结构图

输出层为34个神经元，每个神经元的值为0或者1且只有一个为1，标注为1的那个表示输出结果为那个类别。若正确，则分类成功，若错误，则分类失败。

本文所提出的DBN算法是基于Python的Theano科学计算库，实验环境是4核Intel(R) Core(TM) i5-4460 CPU @ 3.20GHZ,8GB内存 DDR3。训练过程：首先通过预训练对网络参数进行初始化赋值；然后进行调优，逐层训练生成网络的各个参数，达到可以分类的功能。网络由1000*1000*1000个隐藏神经元组成，有Softmax层，RBM层，输入输出层。预训练的学习率是0.01，调优的学习率是0.10。测试车牌字符的汉字准确率达到到了99.44%。相比于神经网络如图10，准确率有了很大的提升。

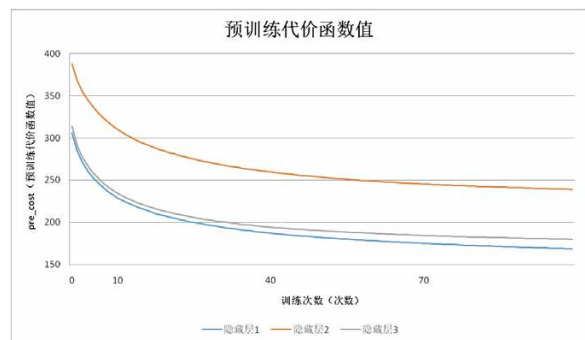


图8 预训练代价函数值变化图

图8中纵坐标为输出结果与标签的平方差的和，所以它的值是越小越好；横坐标是训练次数，可见随着训练次数的增加，深信度网络的代价函数值在减少。通过预训练来减少代价函数值，将使得调优收敛快速且准确。在对网络模型进行参数调优训练时，训练次数与错误识别率的变化如图9所示。

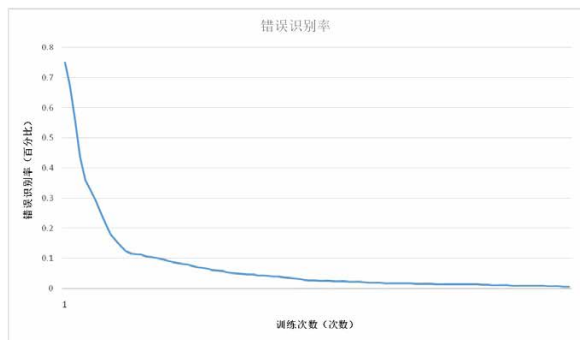


图9 错误识别率随着训练次数增加的变化情况

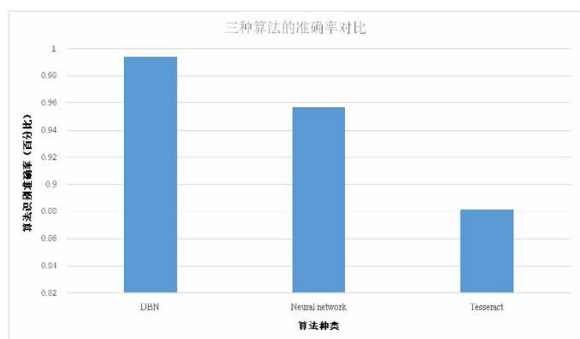


图10 DBN, Neural Network, Tesseract三种算法准确率比较

表1 各种算法在训练时间上的对比情况

算法	神经网络	深信度网络	Tesseract
训练时间(min)	3.36	2.09	2.48

在用深信度网络对车牌汉字进行训练时，随着深信度网络的训练次数的增加，错误识别率在逐渐减少，网络识别准确率逐渐收敛于99.44%。在车牌汉字识别的准确率上，深信度网络算法与神经网络算法，Google的Tesseract算法的对比如图10所示，在训练时间上的对比情况如表1所示。

可以看出，在保证准确率的情况下，深信度网络在所消耗的训练时间比神经网络少的多，与神经网络算法和Tesseract算法相比，深信度网络算法的车牌汉字的准确率要高得多。

深度学习在人工智能领域的应用开始逐渐展开，在快速发展的信息时代，智能家居，智能交通等的发展也跟上了脚步。随着车牌识别技术的逐渐发展，为今后的车辆跟踪，交通监控等领域的应用做好了铺垫。

4. 总结

对于迎接大数据时代的到来，交通领域作为大数据的基础部分，与其相关的大数据技术的研究成为重中之重。本研究分析了一些特征提取算法的优缺点，用RBM构建了一个DBN算法，通过车牌汉字识别方面的应用，验证了所提出算法的准确率可达99.44%。比传统神经网络算法的准确率高得多，而且训练复杂度也大幅降低。

虽然本研究提出的深信度网络车牌字符识别方法有较高的识别率，但仅限于PC机上的实验，对于车牌汉字这类少量种类的训练，完全可以将算法移植到树莓派等小型处理机上。下一步工作将继续改进算法，使算法准确率得到进一步提升的同时，加快训练速度和识别速度，这样有助于将算法移植到便捷式终端，使得智能车库，智能交通成为现实。

参考文献：

- [1] 王笑京, 沈鸿飞, 汪林. 中国智能交通系统发展战略研究[J]. 交通运输系统工程与信息, 2006 (4):9 - 12.
- [2] 吴佳. 车辆牌照识别系统的设计与实现[D].北京：北京交通大学, 2015.
- [3] 公安部. 中华人民共和国公共安全行:中华人民共和国机动车号牌[M]. 中华人民共和国公安部. 2014 - 01 - 24.
- [4] Wu F, Wang Y G, Hou X W. License plate character recognition based on framelet[C]//International Conference on Wavelet Analysis and Pattern Recognition, 2007:673 - 676.
- [5] Cheng R, Bai Y. A Novel Approach for License Plate Slant Correction, Character Segmentation and Chinese Character Recognition[J]. International Journal of Signal Processing Image Processing and Pattern Recognition, 2014, 7(1):353 - 364.
- [6] 汪启伟. 图像直方图特征及其应用研究[D]. 中国科学技术大学, 2014.
- [7] 蔺海峰, 马宇峰, 宋涛. 基于SIFT特征目标跟踪算法研究[J]. 自动化学报, 2010, 36(8):1204 - 1208.
- [8] CIRESAN D U, Masci J, et al. A committee of neural networks for traffic sign classification[C]// IN INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS. 2011:1918 - 1921.
- [9] Google Tessact [EB/OL].[2015 - 10 - 19] <http://sourceforge.net/projects/tesseract-ocr/>
- [10] B Y, C A, V P. Representation learning: a review and new perspectives.[J]. IEEE Transactions on Pattern Analysis & Machine

Intelligence, 2013, 35(8):1798 - 1828.

[11] Bengio Y. Learning Deep Architectures for AI[M]//. Foundations and trends in machine learning, Boston: Now Publishers Inc, 2009, 2(1):1 - 127.

[12] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets.[J]. Neural Computation, 2006, 18(7):1527 - 1554.

[13] Hinton G E. Deep belief networks.[J]. Scholarpedia, 2009, 4(5):5947.

[14] Fischer A, Igel C. Training restricted Boltzmann machines: An introduction[J]. Pattern Recognition, 2014, 47(1):25 - 39.

[15] Le R N, Bengio Y. Representational power of restricted Boltzmann machines and deep belief networks[J]. Neural Computation, 2008, 20(6):1631 - 1649.

[16] 刘忠, 茆诗松. 分组数据的Bayes分析——Gibbs抽样方法[J]. 应用概率统计, 1997, (02):211 - 216.

要闻集锦

Top500出炉：IBM Summit和Sierra超级计算机夺得榜首

2018年11月发布的Top500榜单显示，曾经在6月首次获得第一名的美国能源部Summit超级计算机，再次位列第一巩固了领先优势。这台超级计算机最近升级之后的性能得分达到143.5petaflops，之前为122.3petaflops。

在全球超级计算机排行榜上，美国目前有两台超级计算机领跑，位列第一和第二的分别是IBM打造的两套系统。

最新发布的Top500榜单显示，曾经在6月首次获得第一名的美国能源部Summit超级计算机，再次位列第一巩固了领先优势。这台超级计算机最近升级之后的性能得分达到143.5petaflops，之前为122.3petaflops。

Top500出炉：IBM Summit和Sierra超级计算机夺得榜首

与此同时，Summit的姊妹超级计算机Sierra成为新晋第二名，在升级后取代了中国的“神威太湖一号”（Sunway TaihuLight），性能得分为94.6petaflops。

Top500榜单采用Linpack数学测试来评估计算机执行计算的速度。Summit超级计算机的处理能力相当于世界上每个人每秒执行1900万次计算，这一壮举可以说是令人印象深刻。

另外还采用了另一个名为High - Performance Conjugate Gradient的测试，在这项测试中Summit和Sierra也

排名第一。

超级计算机通常用于研究目的，包括核弹虚拟测试、了解宇宙的形成、预测气候变化、飞机空气动力学建模等任务。

橡树岭国家实验室所拥有的Summit专为高能物理和材料发现等人工智能工作负载设计。该实验室称，在某些情况下Summit每秒可执行超过30亿次的计算。

Sierra则是由美国能源部国家核安全局和劳伦斯利弗莫尔国家实验室共同运营的。

两台超级计算机均由IBM Power9 CPU和Nvidia V100 GPU的组合提供动力，是非常庞大的，由很多冰箱大小的电脑机柜组成。Summit拥有240万个处理器核心，而Sierra拥有160万个核心。

Top500的第三名和第四名均来自中国。“太湖之光”（TaihuLight）超级计算机以93petaflops的表现滑落至第三位，而天河2A则以61.4 petaflops的成绩获得第四。排名第五位的是用于瑞士国家超级计算机中心的Piz Daint，性能得分为21.2 petaflops。

中国打造的超级计算机仍然在整体排名中占据主导地位，前500名中有227套来自中国。与此同时，美国只有109套，历史新低。

（陈继军）