1. 针对具体业务场景或算法的优化

* 【论文】Adaptive Consensus ADMM for Distributed Optimization

【问题】乘法器的交替方向法（ADMM）通常用于分布式模型拟合问题，但其性能和可靠性强烈依赖于用户定义的惩罚参数。理论上，ADMM收敛于任何常数惩罚参数；然而，在实践中，ADMM的效率对这个参数选择非常敏感。

【解决】提出了O（1 / k）自适应ADMM方法的收敛速度节点特定的参数，并提出自适应共识ADMM（ACADMM）方法，它可以通过曲率信息来自动调整参数。

【注释】交替方向乘法器（ADMM）是一种通过将凸分解成更小的且更易处理的部分来解决凸优化问题的算法。

* 【论文】A Communication-Efficient Parallel Algorithm for Decision Tree

【问题】随着大数据的出现，越来越需要并行化决策树的培训过程。然而，大多数沿着这条线路的现有尝试都遭受高昂的通信费用。

【解决】提出了一种称为并行投票决策树（PV-Tree）的新算法，在每次迭代中执行本地投票和全局投票。本地投票根据本地数据从每台机器中选择top-k属性，全局投票通过本地投票的结果再次投票选出top-2k属性。最后，从本地机器收集全局top-2k属性的全粒度直方图，以便确定最佳属性及其分割点。PV-Tree可以实现非常低的通信成本（独立于总数属性数量），因此可以非常好地扩展。

* 【论文】Communication-Optimal Distributed Clustering

【问题】聚类大型数据集是机器学习中的一些应用程序的基本问题。数据通常收集在不同的站点上，我们希望分布式设置中集群的效果与集中式设置中（所有数据驻留在单个站点上）的效果相匹配。

【解决】研究了两个分布式模型中的聚类问题：（1）点对点模型;（2）具有广播频道的模型，并给出了两种模型中的协议，证明显示它们几乎是最优的。对于分布在s个服务器中的图谱中的n个点或n个顶点，在最坏情况下划分点对点模型中的通信复杂度为n·s，而在广播模型中为n +s秒。

* 【论文】Feature-distributed sparse regression: a screen-and-clean approach

【问题】大多数现有的分布式稀疏回归方法假定数据按样本进行分区。 但是，对于高维数据，通过特征划分数据更自然。

【解决】提出了一种分布式稀疏回归算法，当数据通过特征而不是样本进行分割时，允许用户通过折算通讯网络和通讯次数发送的总数据量来调整。 我们表明，我们的方法的实现能够在几分钟内解决数百万个功能的l1l2正则化回归问题。

* 【论文】GPU-based parallel optimization of immune convolutional neural network and embedded system

【问题】在图像识别系统的设计中，传统的卷积神经网络存在一些缺陷如训练时间长，容易过拟合，分类错误率高等。

【解决】运用免疫机制对卷积神经网络进行改进，并提出了一种新的免疫卷积神经网络算法，分析了卷积神经网络的网络结构和参数。算法不但集成了网络节点的位置数据和可调参数，还动态调整了基函数的平滑因子。

【注释】免疫算法是受生物免疫系统的启发而推出的一种新型的智能搜索算法，是一种确定性和随机性选择相结合并具有“生成+检测”能力的启发式随机搜索算法。

* 【论文】Accelerating deep neural network learning for speech recognition on a cluster of GPUs

【问题】我们采用分布式GPU集群的方式训练深度神经网络来解决大词汇量连续语音识别的声学建模问题。并证明ASGD并行化方法对于这个应用程序来说效率不高。即使使用4个GPU，也会导致大量开销而且达不到足够的精确度。

【解决】采用P-learner K-step模型平均算法（KAVG）来明确管理并行实现中的梯度单调性(gradient staleness)。为了扩展超过4个GPU，我们引入了自适应批量大小方案，用以处理非常大的（不适合GPU内存的）批量大小。

* 【论文】A Communication Eicient Parallel DBSCAN Algorithm Based on Parameter Server

【问题】最近的研究表明，基于MPI的DBSCAN分布式实现（例如PDSDBSCAN）优于其他实现，如apache Spark等。但是，MPI DBSCAN的通信成本随着处理器数量的增加而急剧增加，这使得它不适用于大规模问题。

【解决】提出PS-DBSCAN，一种结合了不相交集数据结构和参数服务器框架的并行DBSCAN算法，以最大限度地降低通信成本。由于同一集群内的数据点可能分布在不同的工作人员之间，导致出现多个不相交集合，因此将它们合并会导致巨大的通信成本。在我们的算法中，我们采用快速全局联合方法来结合不相交集以减轻通信负担。

【注释】DBSCAN是一种基于密度的聚类算法，这类密度聚类算法一般假定类别可以通过样本分布的紧密程度决定。在某一类别任意样本周围不远处一定有同类别的样本存在。

　　通过将紧密相连的样本划为一类，这样就得到了一个聚类类别。通过将所有各组紧密相连的样本划为各个不同的类别，则我们就得到了最终的所有聚类类别结果。

* 【论文】Network–Efficient Distributed Word2vec Training System for Large Vocabularies

【问题】Word2vec是一种流行的算法，用于无监督地训练大型文本语料库上的单词的密集向量表示。无论是将所有词汇词的向量存储在单个服务器的存储器中，还是由于大规模网络数据传输而遭受不可接受的训练等待时间，都使得现有的word2vec训练系统对于训练超大规模词汇量的新应用不切实际。

【解决】提出一种新颖的分布式并行训练系统，该系统能够在共享的商品服务器集群上为数亿个词汇的词汇表提供前所未有的矢量实用培训，使用的网络流量远远低于现有解决方案。

* 【论文】TencentBoost: A Gradient Boosting Tree System with Parameter Server

【问题】梯度提升树（GBT）是一种广泛使用的机器学习算法，尽管实现GBT的现有可伸缩系统（如XGBoost和MLlib）对于具有中等维度特征的数据集性能良好，但对于训练数据集包含高维特征的许多工业应用来说，它们可能会遭受性能下降。

【解决】提出了一个使用可扩展的参数服务器体系结构来促进模型聚合，同时引入稀疏拉伸方法和高效的索引结构来提高处理速度。

* 【论文】Reducing view inconsistency by predicting avatars' motion in multi-server distributed virtual environments

【问题】多服务器分布式虚拟环境（DVE）系统包含一组服务器以支持大量地理分布的用户。它提供了一个共享的虚拟场景，其中用户通过替身进行网络交流。然而，在设计具有这种架构的高效DVE系统时存在两个技术挑战：查看由于用户分布不均导致的网络延迟和服务器过载造成的不一致。

【解决】提出了一个新的分区框架来同时解决这两个问题，其中DVE方案被分成一组不相交的网格，称为虚拟单元。然后通过区域四叉树对虚拟单元进行索引，并且提出了一种算法来搜索区域四叉树以便在服务器之间平均分配替身。通过使用概率模型预测每个替身的未来运动和其感兴趣区域将覆盖的虚拟区域。通过估算替身未来位置到每个集群的几何中心的可能距离，提出了一种算法来重新分配服务器之间的未分配替身，以减少视图不一致并保持服务器之间的工作负载平衡。

* 【论文】PR-ELM: Parallel regularized extreme learning machine based on cluster

【问题】极限学习机（ELM）由于其高效和易于实施的特点而在过去的十年中得到了深入的研究。近来，已经提出许多变体，例如并行ELM（PELM），增量ELM和在线顺序ELM（OS-ELM），以改进其时序性能并使其能够增量学习。

【解决】我们提出两个并行变体，称为数据并行正则化ELM（DPR-ELM）和模型并行正则化ELM（MPRELM），以进一步提高ELM在处理大规模学习任务时的计算效率。所提出的PR-ELM算法优于现有变体的优点如下所示：

（1）它们独立地训练每个数据块或每个子模型，因此它们具有更好的并行性。

（2）整个数据集或整个模型被分割成小块或子模型，因此它们大大减少了大量运行时内存的需求。

（3）它们可以在具有更多计算节点的群集上进行配置，因此DPR-ELM和MPR-ELM都具有更好的可扩展性。

* 【论文】Update schedules for improving consistency in multi-server distributed virtual environments

【问题】分布式虚拟环境（DVE）已广泛应用于在线游戏，军事训练和协作工程应用等。但是，由于几个原因，服务器仍然可能会饱和网络带宽。在这种情况下，状态更新不能及时传播，虚拟世界的一致性无法得到保证。

【解决】首先，我们在理想情况下分析DVE，假设DVE的配置保持不变。 我们制定了寻找最佳更新时间表的问题，以便将总TSI最小化为一个不等式约束问题（ICP）。

其次，使用拉格朗日乘子来帮助推导可用于实际系统的启发式更新算法。

* 【论文】Fast ADMM Algorithm for Distributed Optimization with Adaptive Penalty

【算法】提出了一种新的方法来加速交替方向乘法器（ADMM）的交汇，通过在每次迭代中自动决定参数一致所需的约束惩罚来加速收敛速度。另外，我们还提出了自适应确定更新惩罚的最大迭代次数的方法的扩展，并证明这种方法有效地导致了分布式优化的自适应动态网络拓扑结构。

1. 算法分析

* 【论文】Failures of Gradient-Based Deep Learning

【算法】描述了四种深度学习中的简单问题，在这四种问题中，基于梯度的深度学习会失败或者遇到极大阻碍。

* 【论文】An optimized second order stochastic learning algorithm for neural network training

【算法】提出了一种改进的随机二阶学习算法用于监督神经网络训练，称为有界随机对角线Levenberg-Marquardt（BSDLM），它利用梯度和曲率信息来实现快速收敛，而只需要最小的计算开销而不是随机梯度下降（SGD）方法。 B-SDLM只有一个超参数，因此不会受到超参数过拟合问题的困扰。使用多层感知器（MLP）和卷积神经网络（CNN）模型的实验表明，B-SDLM在分类准确性和计算效率方面优于其他学习算法。

* 【论文】Asynchronous Stochastic Gradient Descent with Delay Compensation

【问题】异步随机梯度下降（Asynchronous Stochastic Gradient Descent，ASGD）被广泛用于使用大量训练数据来学习大型神经网络，但是，它仍然受到梯度延迟问题的困扰。也就是说，当本地worker将梯度添加到全局模型时，全局模型可能已被其他worker更新，从而使得梯度产生延迟。

【解决】提出一种通过利用梯度函数的Taylor展开和Hessian的有效逼近来实现的损失函数矩阵的算法，称之为延迟补偿ASGD（DCASGD）。对CIFAR-10和ImageNet数据集上的算法进行的评估表示， DC-ASGD优于同步SGD和异步SGD，且几乎接近顺序SGD的性能。

* 【论文】On Distributed Deep Network for Processing Large-Scale Sets of Complex Data

【问题】随机梯度下降（SGD）可能是训练深度神经网络最常用的优化程序。然而SGD的传统表述本质上是连续的，使得将它应用于非常大的数据集上不切实际。 如果一台机器出现故障，则整个训练过程的时长将取决于最慢的机器。

【算法】提出了一种异步随机梯度下降过程中有效的方法：Bagging-Down SGD，提高学习速率并支持大量模型副本，以确保在处理大规模数据时的学习速率和准确性。

* 【论文】Asynchronous Distributed Semi-Stochastic Gradient Optimization

【问题】现有的随机梯度下降（SGD）算法由于随机梯度的固有方差，很多收敛缓慢，一些收敛比较快的方法也是以牺牲准确率为代价的。

【解决】提出了一种快速分布的基于SGD的异步减少算法。 可以使用恒定的学习速率，并且还可以保证线性地收敛到最佳解决方案。

* 【论文】Weight Normalization: A Simple Reparameterization to Accelerate Training of Deep Neural Networks

【问题】基于一阶梯度的优化高度依赖于被优化的目标曲率。能有多种参数化同一模型的等价方法，寻找神经网络参数化的好方法是深度学习中的一个重要问题。

【解决】提出一种权重归一化的方法，用于提高神经网络模型权重的优化性，加速随机梯度下降优化的收敛。

* 【论文】A Comprehensive Linear Speedup Analysis for Asynchronous Stochastic Parallel Optimization from Zeroth-Order to First-Order

【算法】本文通过全面而通用的分析，研究了从零阶到一阶法的各种异步并行随机算法的加速性能，并首次提出了一种新颖的异步并行零阶方法。

* 【论文】Asynchronous Coordinate Descent under More Realistic Assumption

【算法】证明了有界和无界延迟下的异步BCD收敛结果的选择，以及随机和确定性块选择。 到目前为止，这些结果不需要绝大多数的独立性假设。 因此，异步BCD可以更好地模拟真正的异步求解器的行为。