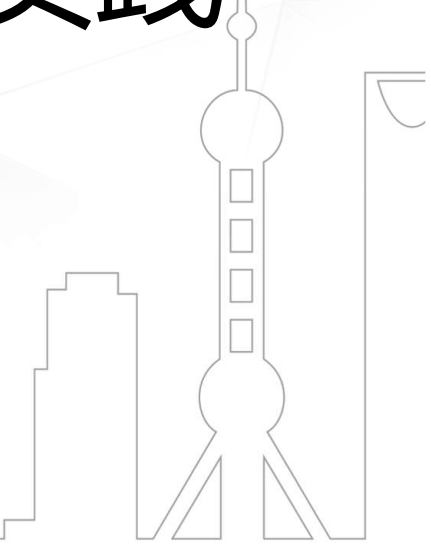
QCon[上海站]
全球软件开发大会2016

深度学习框架的性能优化及其在医药行业的应用实践

SPEAKER

朱 智 勇
英特尔亚太研发中心



International Software
Development Conference

主办方 Geekbang. 极客邦科技 InfoQ

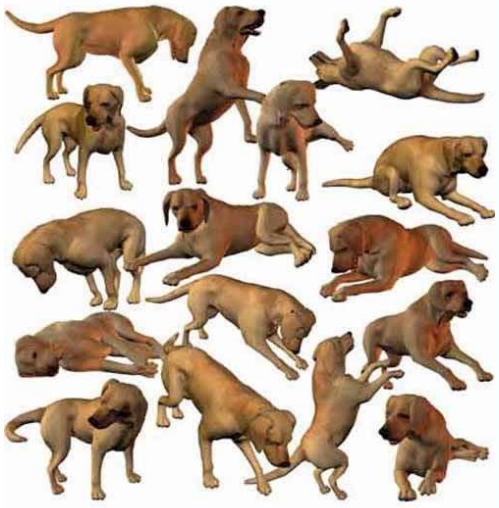
深度学习框架及性能优化

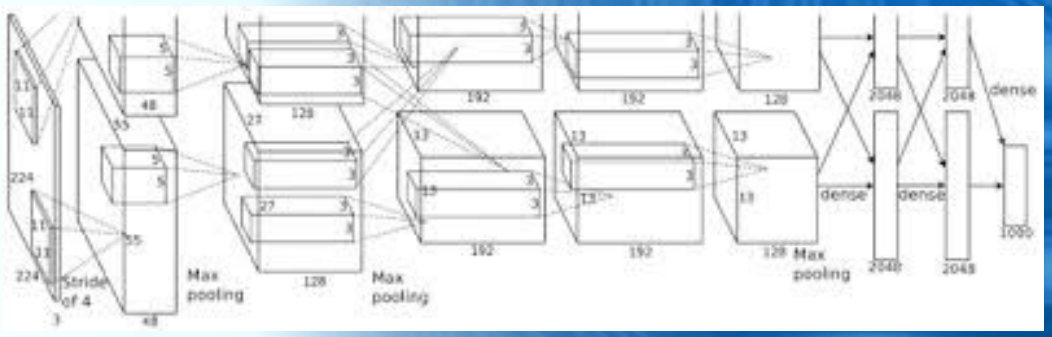
深度学习在医药行业案例分析

深度学习介绍

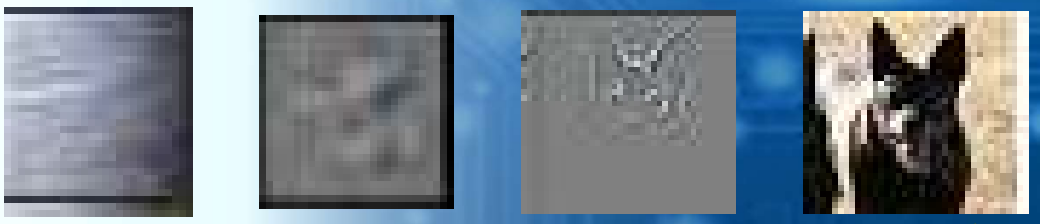
- 机器学习的一种
- 神经网络
- 深层线性和非线性

- CNN/RNN/DBN等模型
- 图形/图像/语音/文本等应用





“dog”



Deeper Layers of the Network → Higher Level Features

深度学习的性能优化

全面的软件优化

涵盖主要深度学习框架

浮点能力的优化

Cache和memory的优化

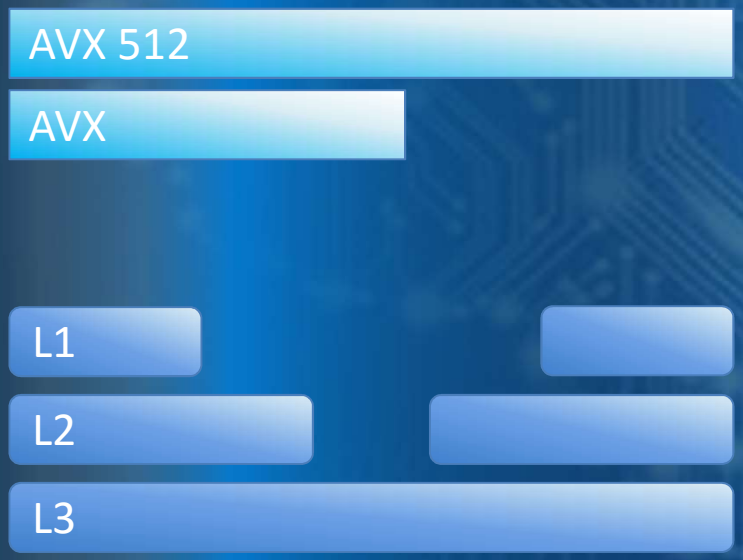
并行计算的支持

丰富的调优软件

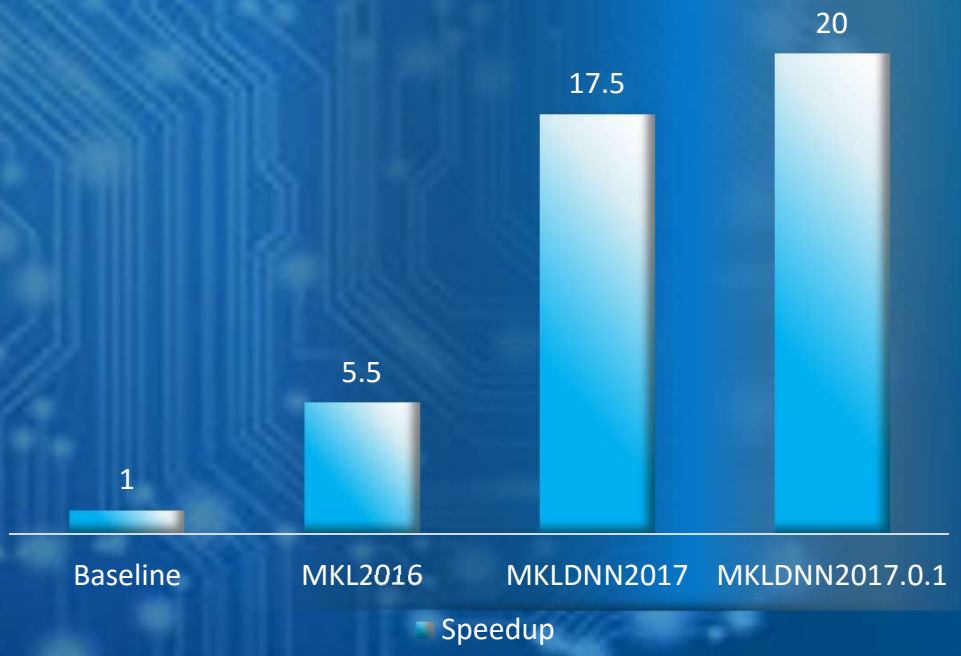
https://software.intel.com/machine-learning/

单节点优化

Single node optimization

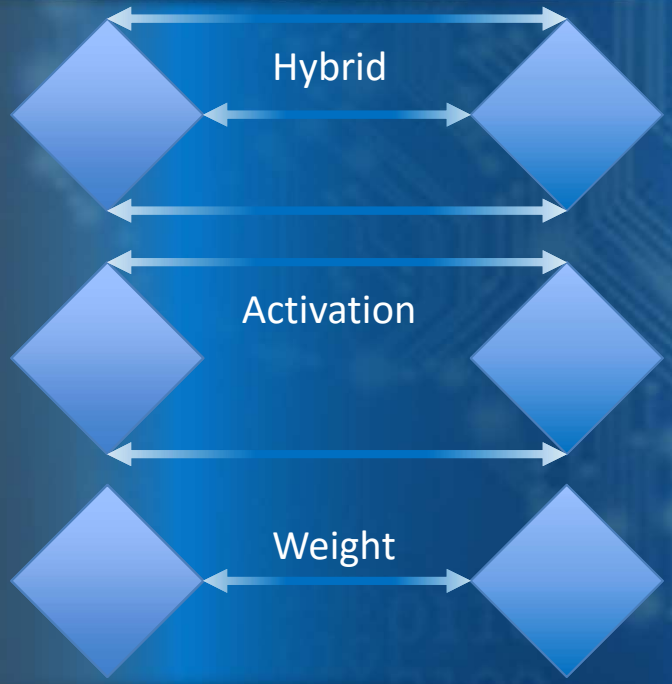


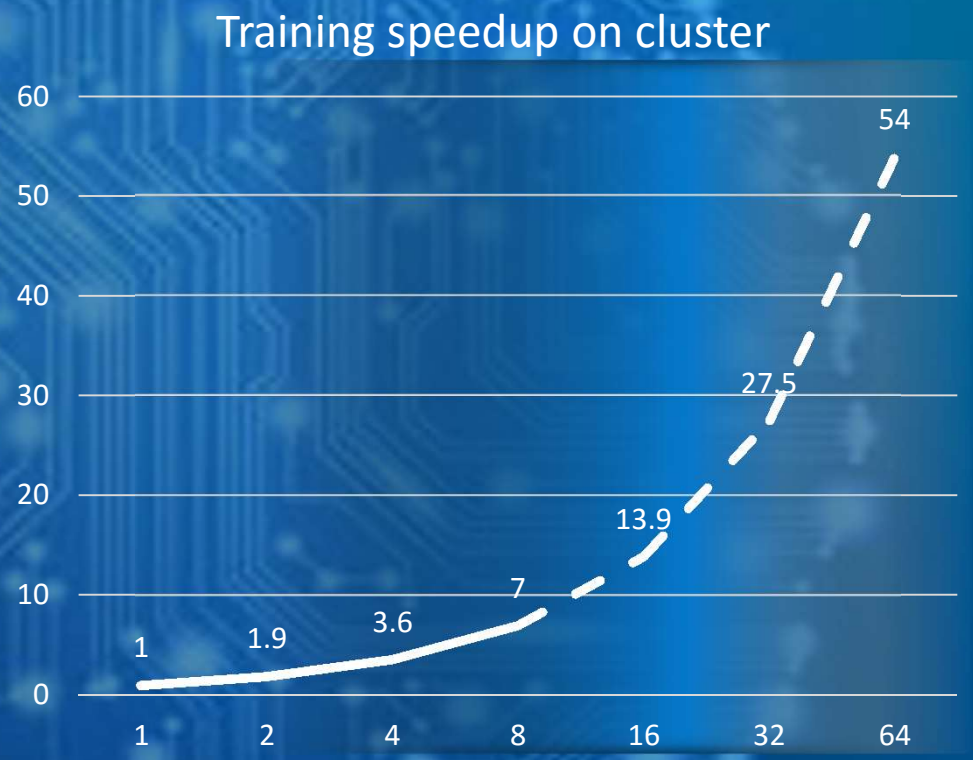
Alexnet training on Xeon BDW



单节点性能不断飞跃

多节点支持





先进的并行技术

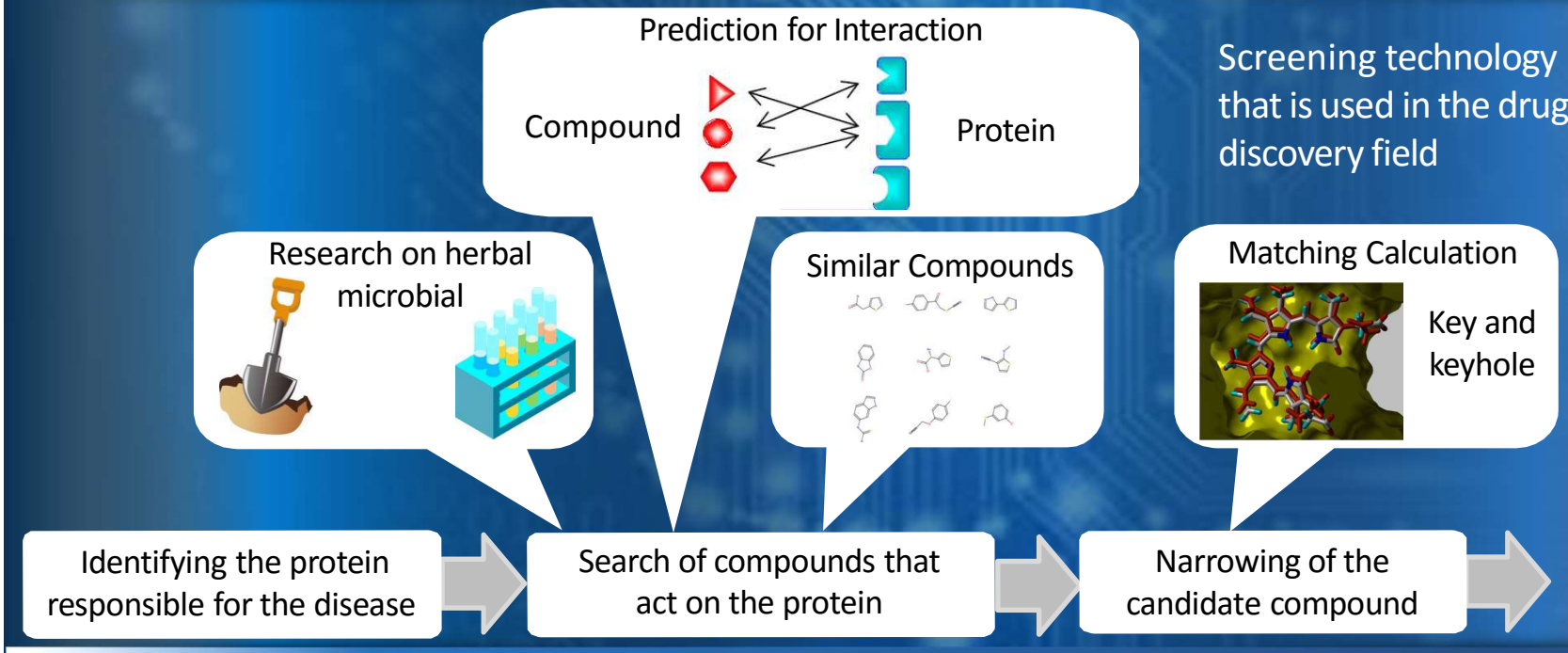
深度学习框架及性能优化

深度学习在医药行业案例分析

案例一：药物研发

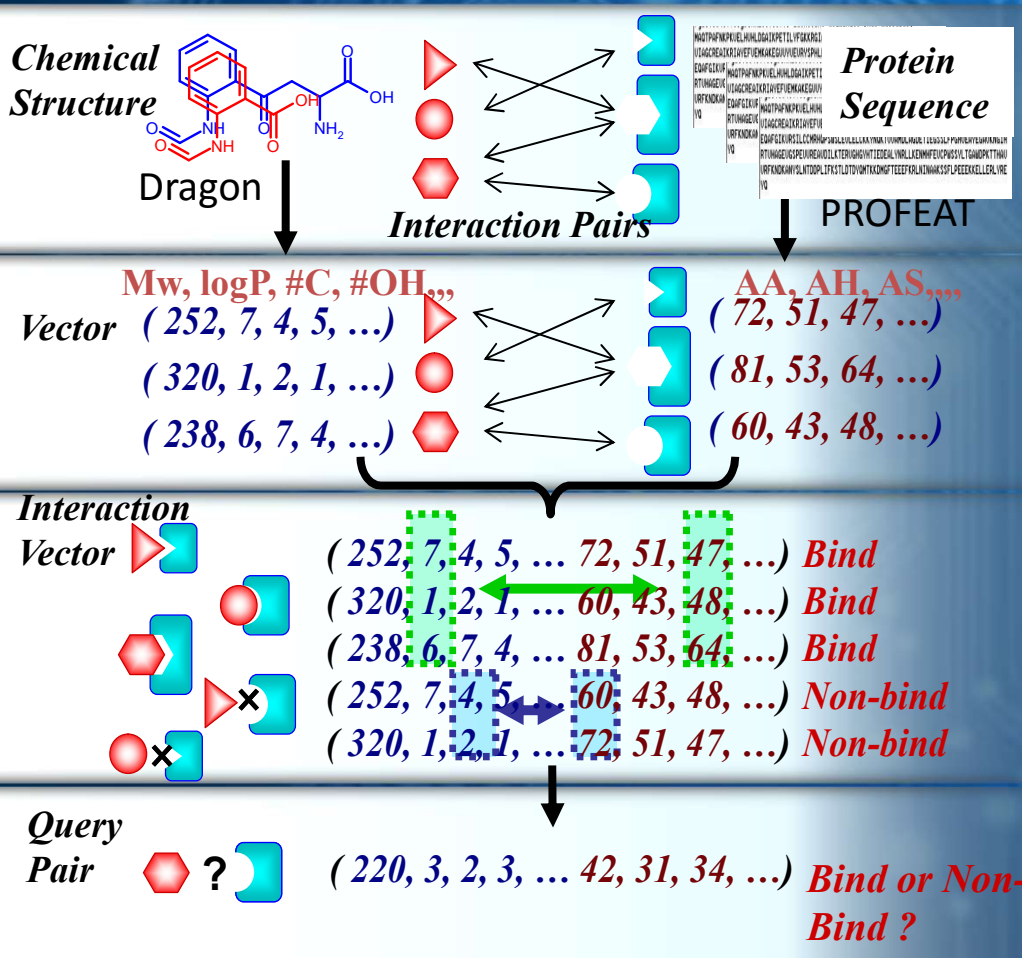
医药研发中的问题：研发周期长，成功率低

用深度模型：筛选候选化合物，缩小研究范围



深度学习在药物研发中的应用

化合物与蛋白质的匹配



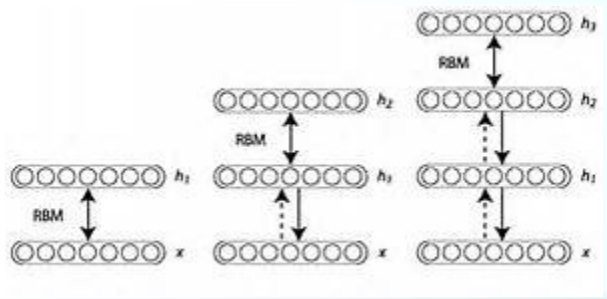
将化合物表达为向量

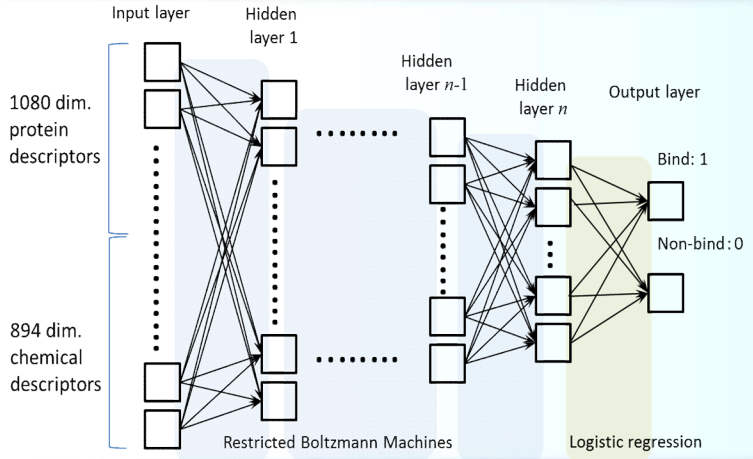
将蛋白质表达为向量

数据：向量的组合

分类：是否有效组合

解决方案：Deep belief network





最终的方案：

- 基于Intel Xeon服务器
- 更多的内存
- 更快的训练速度
- 更多的数据
- 更高的预测精度

更多内存 更快速度 更高精度

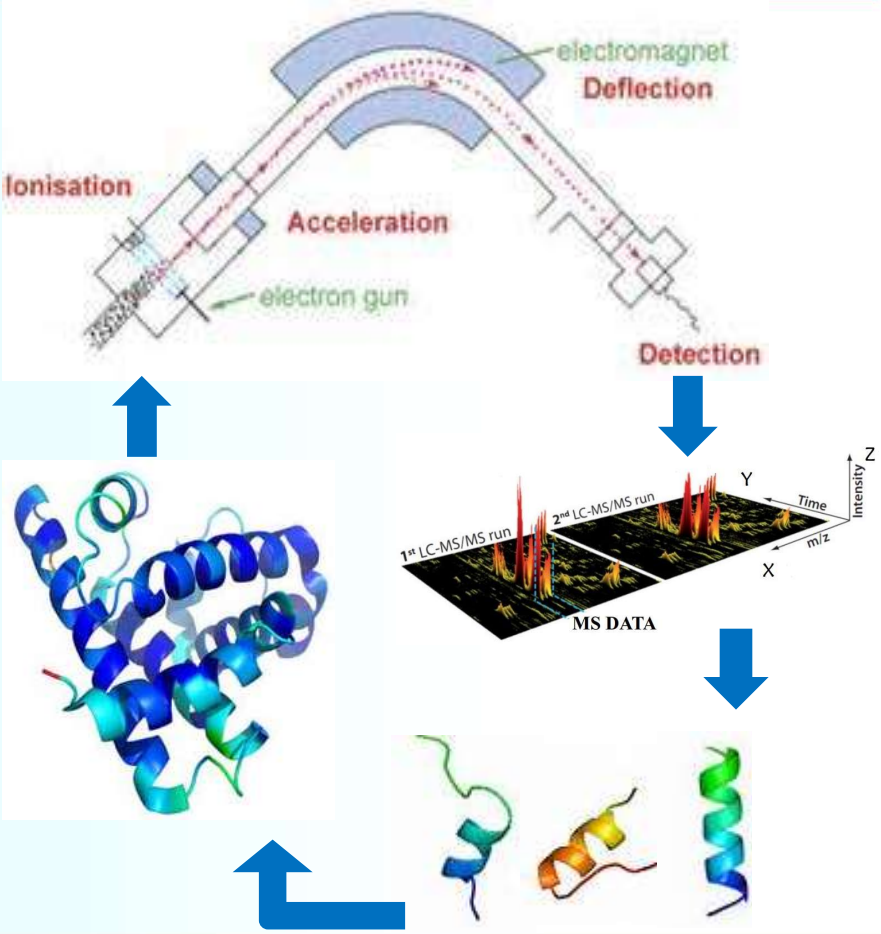
案例二：蛋白质分析

蛋白质组学

- 蛋白质分解
- 双联质谱仪
- 肽链的分析
- 蛋白质的结构，表达和功能

问题建模

- 预测肽链的出峰时间



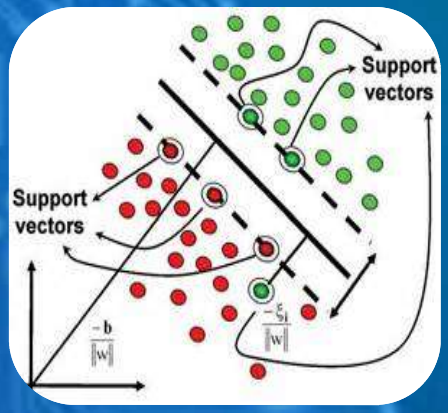
蛋白质组学对于疾病研究有重要意义

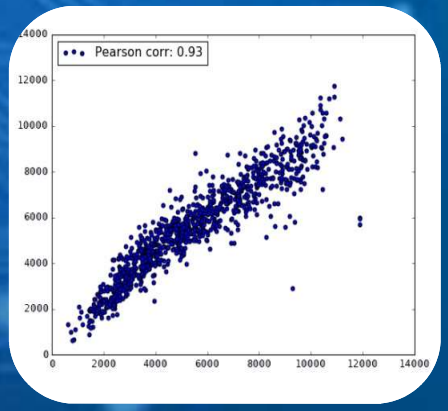
传统的预测方法

传统的预测方法

- 基于SVM
- 时间较长
- 准确率中等

Peptides	RT(min)
SALLALGLK	120.22
NALSSLWGK	122.51
YATLATVSR	53.88
YPMAVGLNK	71.74
NPITNALVR	83.94
QAYTQFGGK	42.86
VYGYVTNSK	??
FVYSLLGPR	??

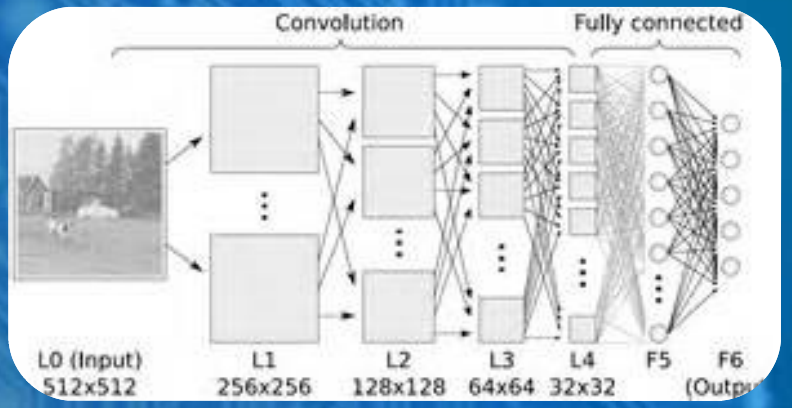


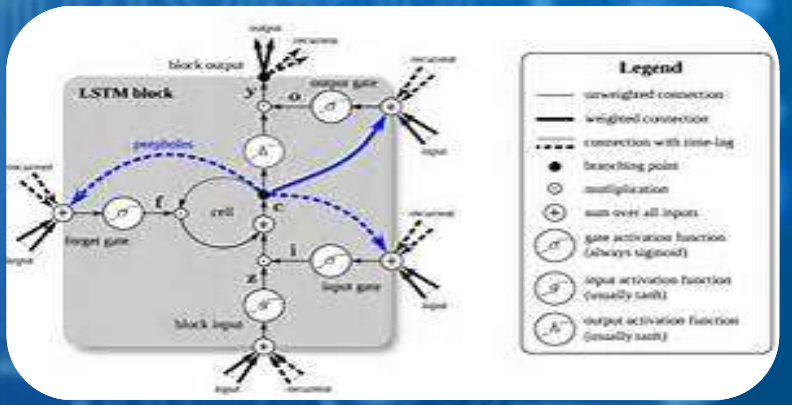


深度学习的方案

深度学习方案

- CNN模型
- LSTM模型
- 1/3训练时间
- 突破的预测准确率



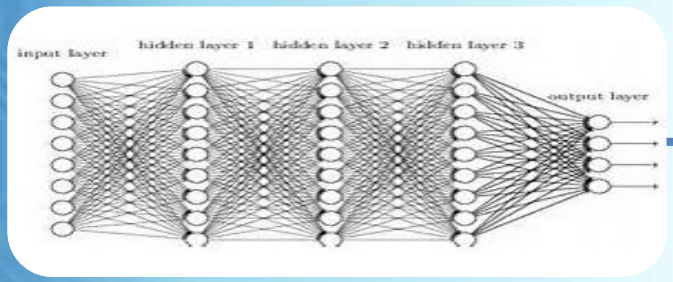


质谱仪 RT 预测的突破

深度学习 + 机器学习

组合方案

- 特征提取：CNN 和 LSTM
- 回归拟合：机器学习方法
- 预测结果：Bagging
- 最好的预测结果



深度学习 + 机器学习 = 更好结果

总结

全面的支持和优化，单节点及分布式

更多内存，更多适用性，更快速度，更高精度

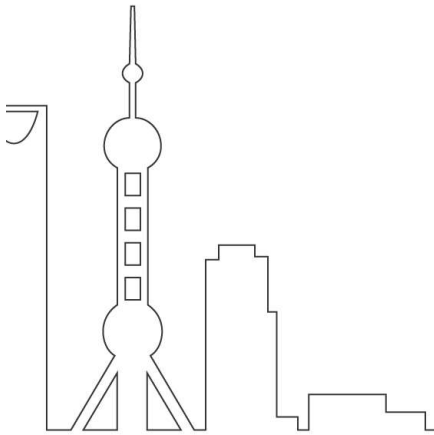
深度学习在医药领域有很好应用，其他行业也有机会

深度学习和机器学习结合，会有更好结果

联系方式

朱智勇(Steve)　　邮件：steve.zhu@intel.com

微信：1256646377



Thanks!

International Software Development Conference

主办方 Geekbang. 极客邦科技 InfoQ