说明文档

一．Batch SOM简介

自组织特征映射模型被称为Self-organizing map，由芬兰学者Teuvo Kohonen于1981年提出。该网络是一个由全互连的神经元阵列形成的无教师自组织自学习网络。Kohonen认为，处于空间中不同区域的神经元有不同的分工，当一个神经网络接受外界输入模式时，将会分为不同的反应区域，各区域对输入模式具有不同的响应特征。自组织神经网络模型能将高维数据投影到低维数组中。这种非线性投影生成的二维特征网络可以帮助分析和检测输入空间的特征。其中，Batch SOM 所有的网络权值在每个训练周期的末尾更新，所以可以应用在并行算法中。

在Batch SOM神经网络中，权重公式为：

分别代表训练周期的开始和结束。

标准高斯邻域方程：

是坐标点，是网络宽度，将在每个周期数据输入时收敛直至最后到聚焦到一个神经元。这表征了SOM神经网络的自组织特性：每个输入向量的出现将调整获胜神经元，并使得邻域神经元更靠近输入向量，最终这些神经元将反馈输入向量的似然分布。

在每次网络获得输入向量时，获胜神经元由以下公式决定：

其中,获胜点为c，

二．算法流程



表1.代码参数

流程图

INPUT

12,20

23,24

45,32

Map

Map

Map

Map OUTPUT

<1, 12\*20\*> <1, 23\*24\*> <1, 45\*32\*>

<2, 12\*20\*> <2, 23\*24\*> <1, 45\*32\*>

<3, 12\*20\*> <3, 23\*24\*> <3, 45\*32\*>

…

<K, 12\*20\*> <K, 23\*24\*> <K, 45\*32\*>

Reduce

Reduce

Reduce

Reduce OUTPUT

<1, 12\*20\*+ 23\*24\*+ 45\*32\*>

<2, 12\*20\*+ 23\*24\*+ 45\*32\*>

<3, 12\*20\*+ 23\*24\*+ 45\*32\*>

…

<K, 12\*20\*+ 23\*24\*+ 45\*32\*>

图1.MapReduce过程

三．算法解析

先将初始神经元读入centerlist容器中，进入map过程。

每行INPUT对应一个map过程。

1.计算与输入向量欧氏距离最近的神经元，获得被激活神经元；

2.根据高斯邻域方程

获得激活神经元周围神经元的增益情况。

先算出分子tempup =,分母tempwidth= ，其中

neighbourhoodSize\*exp(-t/decay)，在每个map过程中时间t加一，达到网络收敛效果；

3.得到= exp(tempUp/tempWidth)；

4.将权重公式的分子和分母送入reduce；

每个神经元对应一个reduce过程。

1. 将每个从map中获得的键值对按照key值分解，获得每个神经元对应的分子和分母
2. 将权重公式中的分子和分母累加，即：；
3. 得到一个神经元经过多个向量刺激的结果，即：；
4. 逐个输出K个神经元至centerlist中。

进行下一次迭代，重新开始mapreduce。