# ALS（Spark）算法在英雄推荐中的简单应用

## 概述

本文拿《王者荣耀》游戏中玩家所拥有的英雄信息（该信息主要保存在tbHeroInfo表中），用Spark中的交替最小二乘算法（ALS：Alternating Least Square），做一个简易的英雄推荐模型，推荐的结果用AUC指标来近似衡量。

## ALS(Alternating Least Square)算法简介

假设有二维矩阵A，横轴表示玩家id（player\_id），纵轴表示英雄（hero\_id），矩阵中的元素表示玩家是否拥有该英雄，或玩家对该英雄的评分值，不管元素的具体含义是什么，这里抽象为玩家对英雄的偏好，用rating来表示，值越大表示玩家对该英雄的偏好越大。矩阵A的维度一般较高，但是其中的元素确是比较稀疏的，甚至是非常稀疏的。如果根据玩家的已知偏好，要为玩家推荐其可能感兴趣的英雄，这时候就要设法补全二维矩阵A。补全的办法是把矩阵用两个低维矩阵的乘积表示，即：

A

其中k << min(m, n)，k值可理解为英雄隐含的种类，X矩阵的含义为玩家对不同种类英雄的偏好矩阵，Y矩阵的含义为英雄与不同种类的英雄的关系矩阵。接下来的问题是要如何寻找X、Y矩阵。ALS采用的方法是最小化重构误差，即：

min L(x, y) =

根据上面的公式，结合一些正则化的方法，控制x向量和y向量的大小，进而控制X矩阵和Y矩阵的大小，使模型尽量保持稳定和简单，得到实际的优化目标如下：

min L(x, y) = +

对上面的公式中的x和y分别求导，得到偏导数为0时x和y的表达式，然后再用最优化方法来迭代求解。得到的x的表达式中含有y变量，得到的y表达式中含有x变量，而在两组变量都未知的情况下，在迭代其中一个变量是要固定另一组变量。比如在计算x时固定y值为上一次迭代得到的y值，在计算y时固定x的取值为上一次迭代中求得的x的取值。这种方法即为交替最小二乘法（ALS）。

## 数据概况

玩家数目： 254w

英雄数目： 52

训练和测试数据1：

对所有的玩家-英雄对进行随机划分，80%为训练数据，20%为测试数据；

训练和测试数据2：

对所有的玩家进行随机划分，80%的玩家为训练数据，20%为测试数据；

## 模型验证方法