

# 基于 GraphX 传球网络的传球质量量化研究

廖彬<sup>1</sup> 张陶<sup>2</sup> 国冰磊<sup>2</sup> 于炯<sup>2</sup> 牛亚锋<sup>3</sup> 张旭光<sup>1</sup> 刘炎<sup>4</sup>

(新疆财经大学统计与信息学院 乌鲁木齐 830012)<sup>1</sup> (新疆大学信息科学与工程学院 乌鲁木齐 830046)<sup>2</sup>  
(重庆大学计算机学院 重庆 400044)<sup>3</sup> (清华大学软件学院 北京 100084)<sup>4</sup>

**摘要** 虽然大数据技术在不断成熟,但它在竞技体育领域的相关应用研究还处于探索阶段。常规篮球统计缺乏对传球数据的记录,更缺乏对传球数据的统计分析、价值挖掘及应用等方面的研究。首先,在 GraphX 基础上将传球数据构建成图,为传球质量的研究奠定基础;其次,提出传球质量评估方法 PESV(Pass Expectation Score Value),相比于传统的助攻数与失误数的比值 ATR(Assist Turnover Ratio),PESV 能更全面地评价球员传球的质量;最后,介绍基于传球网络及传球质量评估方法 PESV 的几种应用场景,包括传球质量对比赛结果的影响分析、基于 PESV 值的传球路线选择,并以华人球员林书豪为例,计算其 2015—2016 赛季的传球得分期望值。

**关键词** 大数据应用,传球网络,GraphX,传球质量量化,球员评价

中图分类号 TP393.09 文献标识码 A DOI 10.11896/j.issn.1002-137X.2017.12.033

## Research on Passing Quality Quantification Based on GraphX Passing Network

LIAO Bin<sup>1</sup> ZHANG Tao<sup>2</sup> GUO Bing-lei<sup>2</sup> YU Jiong<sup>2</sup> NIU Ya-feng<sup>3</sup> ZHANG Xu-guang<sup>1</sup> LIU Yan<sup>4</sup>

(College of Statistics and Information, Xinjiang University of Finance and Economics, Urumqi 830012, China)<sup>1</sup>

(School of Information Science and Engineering, Xinjiang University, Urumqi 830046, China)<sup>2</sup>

(School of Computer Science, Chongqing University, Chongqing 400044, China)<sup>3</sup>

(School of Software, Tsinghua University, Beijing 100084, China)<sup>4</sup>

**Abstract** Although the big data technology continues to mature, relevant application research in the field of competitive sports is still in the exploratory stage. Conventional basketball technical statistics lacks the record of the passing data as well as the statistical analysis, data mining, and application of the passing data. Firstly, based on GraphX, we created the passing network graph, which laid a foundation of future research on passing quality. Secondly, PESV (Pass Expectation Score Value), a method of evaluating the quality of passing basketball, was proposed. Compared with the traditional ATR (Assist Turnover Ratio) that defined the ratio of assists to turnovers, PESV can make a more comprehensive evaluation of the passing quality. Finally, we introduced a few application scenarios of PESV based on passing network, including the analysis of the impact of passing quality on game result, passing route selection based on PESV value, and taking the Chinese player Jeremy Lin as an example to calculate his passing expectation values in the 2015—2016 seasons.

**Keywords** Big data application, Passing network, GraphX, Quantitative pass quality, Player evaluation

## 1 引言

自 2003 年 Google 发表论文公开分布式存储系统 GFS<sup>[1]</sup> (Google File System) 及分布式数据处理模型 MapReduce<sup>[2]</sup> 以来,诸多的大数据计算系统及框架(如 Hadoop, Storm, Spark<sup>[3]</sup>, Pig, Hive, Hbase, Dryad 等)以 MapReduce 为计算模型,并形成了以 MapReduce 为核心的大数据计算生态系统。

随着大数据理论与技术的不断发展,人们逐渐认识到大数据最核心的关注点并不是数据的种类(Variety)及量(Volume),而是大数据的价值(Value);并且,大数据时代也并不是意味着所有应用的数据量都大,实际上大数据也是由一个个小数据集合而成的,正是对小数据的持续采集和融合分析,才有积跬步而致千里的大数据价值能量的爆发。

虽然大数据技术在社交网络、金融、公共安全、医疗等方

到稿日期:2017-01-12 返修日期:2017-04-18 本文受国家自然科学基金项目(61562078, 61262088),新疆维吾尔自治区自然科学基金(2016D01B014)资助。

廖彬(1986—),男,博士,副教授,硕士生导师,CCF 会员,主要研究方向为绿色计算、数据库系统理论及数据挖掘, E-mail: liaobin665@163.com;张陶(1988—),女,博士生,主要研究方向为分布式计算、网格计算;国冰磊 女,博士生,主要研究方向为绿色计算、数据库系统等;于炯(1964—),男,博士,教授,博士生导师,主要研究方向为网络安全、网络与分布式计算;牛亚峰(1990—),男,硕士生,主要研究方向为机器学习;张旭光(1994—),男,硕士生,主要研究方向为大数据计算;刘炎(1990—),男,硕士生,主要研究方向为大数据计算。

面的应用逐渐成熟,但在竞技体育方面的应用还处于探索阶段。其中,大数据技术应用于足球方面的首次亮相是在2014年的巴西世界杯,帮助德国再次捧得大力神杯的“秘密武器”之一,则是来自SAP公司的足球大数据技术解决方案——“Match Insights”。Match Insights能够迅速收集、处理分析球员和球队的技术数据,基于“大数据分析”优化球队配置,提升球队的作战能力,并通过分析对手数据,找到在世界杯比赛中的“制敌”方式。利用大数据分析,德国队教练可以迅速了解当前比赛状况、每个球员的特点和表现、球员的防守范围、对方球队的空挡区等信息。通过这些信息,教练可以更有效地优化配置球员的上场时间、位置及战术等,从而提升球员及球队的整体表现。

而大数据在篮球的应用场景中,已有的常规技术统计方法只针对得分、篮板、助攻、抢断、盖帽、失误、犯规、投篮命中率、出场时间等“小数据”,不仅缺乏对传球数据的统计,而且不能通过数据分析量化球员的传球质量。为了解决以上问题,本文做了如下工作:

(1)用GraphX将传球数据构造为一张传球网络图,为球员的传球决策优化及传球质量的量化研究打下基础。

(2)提出传球质量的量化方法PESV(Pass Expectation Score Value),为改进传球质量提供量化依据。

(3)介绍传球网络及传球质量模型的应用场景,包括传球质量对比赛结果的影响分析以及基于PESV值的传球路线选择,并以华人球员林书豪为例,计算其2015—2016赛季的传球得分期望值。

本文第2节对相关工作及背景进行介绍;第3节对传球网络图的构建及可视化进行研究;第4节对传球质量的量化进行研究;第5节介绍几种基于传球网络及传球质量模型的应用场景;最后总结全文,并指出下一步工作的方向。

## 2 相关研究及背景

本文对传球数据的汇聚形态进行研究后发现,大量传球数据的最终汇聚形态为图。而图广泛应用于社交网络(包括用户网络的社区发现、用户影响力、能量传播、标签传播等),用以提升用户的黏性和活跃度;而应用到推荐领域(如标签推理、人群划分、年龄段预测时),则可以提升推荐的丰富度和准确性。除社交网络外,图还在生物信息学、化学信息学、智能交通、舆情监控等领域发挥着巨大的作用。

本文基于GraphX将传球网络构建为图。GraphX基于Spark并扩展了弹性分布式数据集<sup>[4]</sup>(Resilient Distributed Datasets, Spark RDD)的抽象,提出了点和变都带属性的有向多重图<sup>[5]</sup>(Resilient Distributed Property Graph)。GraphX提供Table与Graph两种视图,这两种视图都拥有自己独立的操作符,这使得GraphX的操作具有更大的灵活性。GraphX不仅对Pregel及GAS模型进行了改进,而且其性能也优于GraphLab<sup>[6]</sup>和Giraph<sup>[7]</sup>。GraphX不仅提供节点度(出度/入度)的计算、子图查询、PageRank、最大连通图及最短路径等基本图算法,而且能够无缝地调用SparkCore中的API接口(包括:map, filter, flatMap, sample, groupByKey, reduce-

ByKey, union, join, cogroup, mapValues, sort, partitionBy等)。虽然当前GraphX的应用并不成熟,但是从运行效率及编程模型支撑的角度考虑,运用GraphX进行传球网络数据的构建及分析,是一个较为理想的选择。由于GraphX于Spark版本1.2之后才发布正式版,导致基于GraphX的算法及应用方面的研究工作较少;并且其应用场景通常在互联网领域,如淘宝基于GraphX搭建了图谱平台,可以支撑多图合并、能量传播模型计算、用户影响力计算、商品推荐等算法。在国内,严玉良等人<sup>[8]</sup>提出了一种基于GraphX的大规模频繁子图挖掘算法FSMBUS,相较于GRAMI,该算法的性能提高了一个量级。

在篮球数据研究方面,文献[9]提出了EPV(Expected Possession Value)的概念,将球员场上的行为(如传球、运球、投篮等)转化为动态的EPV值,以量化球员在球场上每次动作的价值。文献[10]关注篮板球数据,将每次篮板数据的产生分解为Positioning, Hustle, Conversion 3个阶段,并基于篮板球数据对NBA中球员的篮板球能力进行了分析。文献[11]对球员加速行为进行了量化研究,并对球员的加速数据进行了可视化。文献[12]对NBA中的投篮选择(2分球与3分球)与对应的风险进行了研究,分析了不同场景下的投篮选择对比较结果的影响。文献[13]提出了通过防守矩阵(Defensive Metrics)的分析来量化球员防守有效性的方法。文献[14]对球员在投篮结束后的冲抢前场篮板与退位防守的选择性问题进行了研究,研究结果表明,退位防守能够提高防守成功率,但减少了抢下前场篮板球的数量。文中还针对不同球队的数据进行了分析,并提出相应的改进策略。

与已有研究工作不同的是,本文首先针对常规的篮球比赛统计中缺乏对传球数据的统计问题,将球员之间的传球关系进行关联(将传球人与接球人作为顶点,顶点之间的传球关系作为边),发现随着传球数据量的不断增加,最终形成了一张稳定的传球网络图;并基于此研究了传球质量的量化算法,为球员的传球质量的量化和分析等应用打下基础。相比于传统的助攻数与失误数的比值ATR(Assist Turnover Ratio),本文提出的传球质量评估方法PESV能更全面地评价球员传球的质量。

## 3 传球网络图的构建及可视化

### 3.1 传球网络图的总体构建流程

传球网络图的构建及应用框架流程图如图1所示,主要分为数据获取、数据清洗、图顶点及边表的设计、传球网络构造4个步骤。其中数据获取方面,一般可通过NBA官方网站或具体场次的录像分析提取数据。通过程序抓取而来的原始数据一般为XML格式或JSON格式,需对其进行数据清洗及格式转化,以满足网络图的构建对数据格式的要求。构建Vertex表及Edge表是构建传球网络图的基础,需根据后期应用的需求确定Vertex表及Edge表中的属性。通过GraphX构建出传球网络图,通过提取并计算图相关的数据(如传球次数、PR值等),结合大数据可视化技术,能够实现传球网络图的可视化。

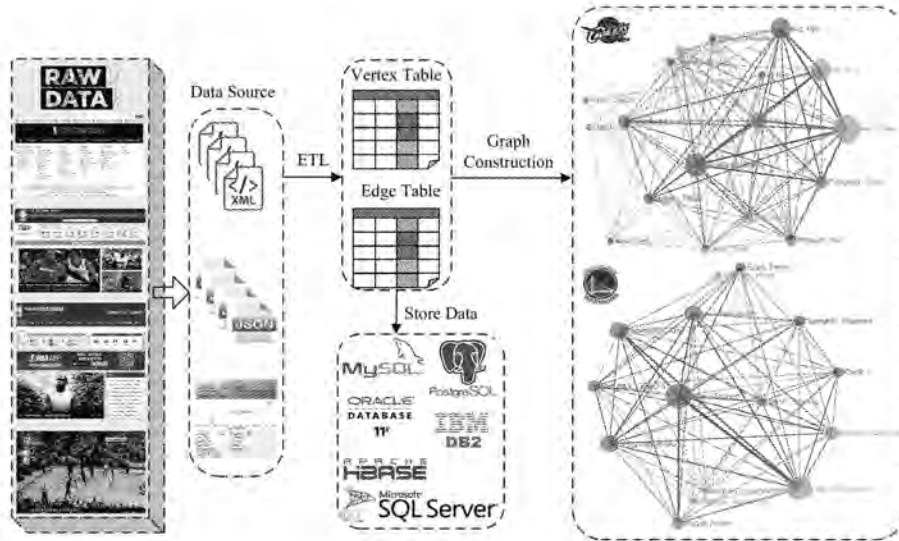


图 1 传球网络图的构建及应用框架流程图

### 3.2 传球数据的获取、清洗及格式转化

NBA 作为世界上顶级的篮球联盟,其官方网站(www.nba.com)于 2016 年开始公开包括传球、防守影响、移动速度及距离等非常规技术的统计。但是,CBA 及中国国家篮球队,对传球数据的获取、统计分析及围绕传球数据的战术、训练等方面的运用还是空白。NBA 官方网站虽然提供传球数据的查询功能,但是并没有提供转存功能,通过手工收集、整理数据的效率较低。图 2 给出通过数据下载程序从 NBA 官网上获取金州勇士队传球数据的流程。首先,收集金州勇士队中每个队员在 NBA 数据库中所对应的 playerid,如:Stephen Curry 的 playerid=201939,Klay Thompson 的 playerid=202691;其次,将收集到的 playerid 组合形成 playerids 数组,该数组中存放着勇士队中所有队员的 playerid;最后,通过调用 CURL API 功能,基于拼接 URL (http://stats.nba.com/stats/playerdashptpass),利用 HTTP Get 方法得到返回的数据,生成当前 playerid 所对应的 JSON 文件。

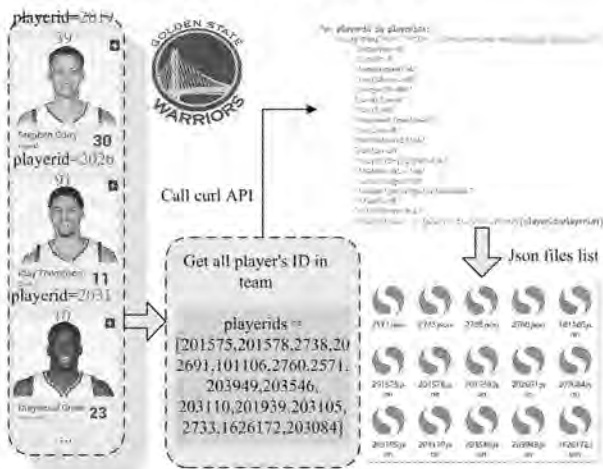


图 2 数据获取流程

### 3.3 定义并构建 Vertex 表及 Edge 表

GraphX 统一了 Table View 与 Graph View,即实现了 Unified Representation。通过构造 Vertex 及 Edge 表,并将

Vertex Edge 表数据加载到 Spark 内存中,便可构造出传球网络图。

**定义 1(传球网络图)** 传球网络图用  $G=(V,E)$  表示,其中  $V=\{p_1, p_2, p_3, \dots, p_n\}$  表示传球网络中所有球员的集合(设传球网络中球员的数量为  $n$ );而  $E=\{\langle p_i \rightarrow p_j \rangle \mid p_i \in V, p_j \in V\}$  表示任意顶点(球员)之间的传球关系的集合,例如: $p_i \rightarrow p_j$  表示球员  $p_i$  传球给  $p_j$ 。

在实际的传球网络应用场景中,设关系表 Vertex (PLAYERID, PLAYERNAME) 及 Edge (PASSERID, PASSTOID, FREQUENCY, PASSNUM, AST, FGM, FGA, FGP, FGM2, FGA2, FGP2, FGM3, FGA3, FGP3) 中分布存储节点集合  $V$  及关系集合  $E$  的数据。Vertex 及 Edge 表数据字典如表 1 和表 2 所列。

表 1 Vertex 表数据字典

字段名	数据类型	意义解释
PLAYERID	INT	球员在 NBA.com 数据库中的 ID
PLAYERNAME	VARCHAR	球员姓名

表 2 Edge 表数据字典

字段名	数据类型	意义解释
PASSERID	INT	传球人 ID
PASSTOID	INT	接球人 ID
FREQUENCY	FLOAT	任意两个球员之间的传球频率
PASSNUM	INT	任意两个球员之间的传球数量
AST	INT	任意两个球员之间的传球助攻数
FGM	INT	接球人接到传球后,命中投篮的数量(Field Goals Made)
FGA	INT	接球人接到传球后,进行投篮的次数(Field Goals Attempted)
FGP	FLOAT	接球人接到传球后的投篮命中率(Field Goal Percentage)
FGM2	INT	接球人接到传球后,2 分球的投篮命中数
FGA2	INT	接球人接到传球后,2 分球的投篮的次数
FGP2	FLOAT	2 分球投篮命中率
FGM3	INT	接球人接到传球后,3 分球的投篮命中数
FGA3	INT	接球人接到传球后,3 分球的投篮的次数
FGP3	FLOAT	3 分球投篮命中率

结合 Spark 中对 Vertices 类的定义,表 1 中字段 PLAYERID 为顶点 ID,字段 PLAYERNAME 为顶点属性。

结合 Spark 中对 Edges 类的定义,表 2 中字段 PASSERID 为源顶点 ID,字段 PASSTOID 为目标顶点 ID,字段 FREQUENCY, PASSNUM, AST, FGM, FGA, FGP, FGM2, FGA2, FGP2, FGM3, FGA3, FGP3 都为边属性。特别地,在 Edge 表中,对于任意的 PASSERID,都需满足约束(设 Edge 表中的 FREQUENCY 字段为变量 *Frequency*):

$$\sum Frequency=1 \quad (1)$$

并且,对于给定的  $\langle PASSERID, PASSTOID \rangle$ , *Frequency* 的计算公式为(设 Edge 表中 PASSERID 与 PASSTOID 字段为变量 *passerid* 与 *passtoid*):

$$Frequency\langle passerid, passtoid \rangle = \frac{passnum\langle passerid, passtoid \rangle}{\sum passnum\langle passerid \rangle} \quad (2)$$

式(2)能计算出任意两个球员之间的传球概率。设 Edge 表中的 FGM, FGA 及 FGP 字段分别为变量 *fgm*, *fga* 及 *fgp*,那么对于给定的  $\langle PASSERID, PASSTOID \rangle$ , *fgp* 可由式(3)计算:

$$fgp = \frac{fgm}{fga} \times 100\% \quad (3)$$

同样地,对于给定的  $\langle PASSERID, PASSTOID \rangle$ , 字段 FGP2 及 FGP3 的值可由式(4)与式(5)计算:

$$fgp2 = \frac{fgm2}{fga2} \times 100\% \quad (4)$$

$$fgp3 = \frac{fgm3}{fga3} \times 100\% \quad (5)$$

在实际比赛中,变量 *Frequency*, *fgp*, *fgp2*, *fgp3* 将对球员的传球决策产生影响。

### 3.4 传球网络的构建

本节以 2015—2016 赛季的总冠军骑士队为例,构建其 82 场常规赛期间所有场次的传球网络图。首先,通过数据获取及清洗获得传球网络图的顶点数据为球队中所有的球员。其中,以球员 LeBron James 为例,其在传球网络中的边数据如表 3 所列。

表 3 LeBron James 常规赛传球网络的 Edge 数据

srcId	dstId	attr
2544	201567	21.5%,820,126,146,322,45.3,80,163,49.1,66,159,41.5
2544	202618	18.9%,721,41,80,191,41.9,64,130,49.2,16,61,26.2
2544	2747	16.6%,634,105,129,313,41.2,42,109,38.5,87,204,42.6
2544	203521	14.6%,558,49,59,124,47.6,22,53,41.5,37,71,52.1
2544	202389	5.7%,216,52,52,102,61.0,52,97,53.6,0,5,0,0
2544	202684	5.5%,209,46,47,76,61.8,47,76,61.8,0,0,0,0
2544	2590	5.2%,199,15,22,51,43.1,15,25,60.0,7,26,26.9
2544	202697	3.9%,148,19,22,61,36.1,12,25,48.0,10,36,27.8
2544	2210	2.9%,109,23,25,50,50.0,18,26,69.2,7,24,29.2
2544	101112	1.7%,64,18,17,32,53.1,2,5,40.0,15,27,55.6
2544	2760	1.3%,48,6,6,15,40.0,6,15,40.0,0,0,0,0
2544	203099	1.2%,47,9,9,17,52.9,3,6,50.0,6,11,54.5
2544	2592	1.0%,40,4,4,17,23.5,0,0,0,0,4,17,23.5

表 3 中字段 srcId, dstId 及 attr 分别对应于 GraphX Edges 类中的源顶点的 ID、目标顶点的 ID 及边属性。其中,attr

可进一步分解为: FREQUENCY, PASSNUM, AST, FGM, FGA, FGP, FGM2, FGA2, FGP2, FGM3, FGA3, FGP3。表 3 只是 LeBron James 单个球员的数据,将骑士队所有队员的传球数据累加,便形成了整个球队 82 场常规赛所有的传球数据。构建常规赛骑士队传球网络的 Spark GraphX 代码如下:

```

1. package Cavaliers, pass, analysis, datamaker
2. import org.apache.spark._
   import org.apache.spark.graphx._
   import org.apache.spark.rdd.RDD
3. object CavalierPassNetWork {
4.   def main(args: Array[String]): Unit = {
5.     val conf = new SparkConf().
       setAppName("CavalierPassNetWork"). setMaster("local
       [4]")
6.     val passes: RDD[String] = sc.textFile("/Cavaliers/edges.
       cvs")
7.     val player: RDD[String] = sc.textFile("/Cavaliers/vertex.
       cvs")
8.     val vertices: RDD[(VertexId, String)] = player.map {
       line => val fields = line.split(" ")
       (fields(0), toLong, fields(1))
       }
9.     val edges: RDD[Edge[(String, Long, Long)]] = pass.map {
       line => val fields = line.split(" ")
       Edge(fields(0). toLong, fields(1). toLong, (fields(2). toS-
       tring(), fields(3). toLong, fields(4). toLong))
       }
10.    val graph = Graph(vertices, edges)
11.  }
12. }

```

在如上传球网络构建代码中,第 6—7 行代码分别将顶点与边的数据从 cvs 文件中读取到内存;第 8 行从 player 常量构建图顶点;第 9 行在第 6 行常量 pass 的基础上构建传球网络的边,边的属性可根据应用场景进行自定义;第 10 行在常量 vertices 及 edges 的基础上,通过调用 Graph 类的构造函数,构造出传球网络 graph,为传球网络的应用打下基础。

### 3.5 传球网络的可视化

大数据可视化技术将枯燥的海量数据通过图表的形式展示出来,能够更加直观地向用户展示数据之间的联系,从而减少用户挖掘数据内涵所耗费的时间。对于具有海量节点和边的大规模网络,如何在有限的屏幕空间中进行可视化,是大数据时代的可视化技术面临的难点和重点<sup>[15]</sup>。Herman 等人<sup>[16]</sup>对图的可视化基本方法和技术进行了综述,虽然研究人员针对图的可视化提出了不少新方法与技术(如树图技术 Treemaps<sup>[17]</sup>、Voronoi 图填充<sup>[18]</sup>、综合性的 TreeNetViz<sup>[19]</sup>等),但是经典的基于节点和边的可视化方法依然是图可视化的主要形式。

传球网络的可视化与社交网络的可视化比较类似,最大的区别是传球网络需要区分球员的重要性、球员角色、传球数据等信息。以球员重要度为例,计算球员重要度(即传球网络中顶点半径的大小)本质上与计算网页重要度算法 PageRank 的思想非常类似,都是将图中节点的重要性映射为一个具体的数字。因此,借鉴 Google 的 PageRank 算法,本文提出计算

球员在传球网络中重要程度的算法——PlayerRank,其计算公式如式(6)所示:

$$PR(p) = (1 - \alpha) + \alpha \sum_{i=1}^n \frac{PR(T_i)}{C(T_i)} \quad (6)$$

其中,  $PR(p)$  表示一个球员  $p$  在传球网络中的重要程度,其取值范围为  $[0, 10]$ , 值越大说明重要性越强;  $\alpha$  为随机权重变量,其取值范围为  $[0, 1]$ , 通常在 PageRank 中  $\alpha$  默认取值 0.85;  $n$  表示传球网络中球员(顶点)的数量;  $T_i (i=1, 2, 3, \dots, n)$  表示向球员  $p$  传球的球员集合;  $C(T_i)$  表示球员  $T_i$  向外传球球员的数目; 而  $\frac{PR(T_i)}{C(T_i)}$  则表示给球员  $p$  传球的球员  $T_i$  给球员  $p$  所赋予的  $PR$  值。

球员角色(如 Guard, Forward, Center, Forward-Guard 及 Forward-Center)在传球网络中用不同的圆表示,任意两个球员之间的传球次数则用不同粗细的边表示。本文从超过 10 种(如 iCharts, Fusion Charts Suit, Modest Maps, Chartkick, Bonsai, Google Charts, Gephi, Protvis 等)可视化工具(框架)中挑选出 networkD3(R 语言中基于 JavaScript 的数据可视化工具)框架中的 forceNetwork 作为传球网络的可视化工具。最终,骑士队传球网络可视化的结果如图 3 所示。

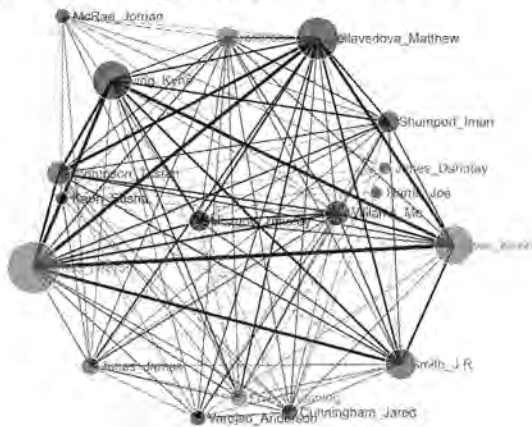


图 3 骑士队常规赛传球网络图<sup>[20]</sup>

## 4 传球质量的量化

### 4.1 单次传球的传球得分期望值

已有的技术统计(如得分、篮板、助攻、抢断、盖帽、失误、犯规、投篮命中率、出场时间等)并不能说明一个球员传球的能力,更不能量化一个球员在传出球时对传球选择的正确性;并且篮球比赛中球员之间发生最频繁的动作就是传球,表 4 所列为 NBA 最近 3 个赛季的总体传球情况。因此,本节将对传球的质量模型进行研究。

表 4 NBA 3 个赛季的总体传球数据统计结果

赛季	比赛总场次	总时间/min	传球总数	场均传球次数	总助攻数	助攻率/%
2013—2014	1226	593280	735553	299.98	53950	7.335
2014—2015	1227	593760	739541	301.36	54069	7.311
2015—2016	1228	593840	747761	304.46	54728	7.319

球员接到传球后,下一步可能发生的动作有运球、投篮及传球 3 种情况,本文主要关注传球行为。发起传球的目的是得分或者让得分更加容易,那么可以用对得分的有利程度来

度量每次传球的质量,其值定义为  $PESV$ (Pass Expectation Score Value),表示一次传球带来的得分的可能性期望。设当球员  $p$  传球给球员  $r$  时,传球得分期望值  $PESV$  如式(7)所示:

$$PESV_{p \rightarrow r} = \begin{cases} -1.0765, & \text{turnover} = \text{true} \\ 2 \times r_{f_{sp2}} \times (1 - da), & \text{2 point shot} \\ 3 \times r_{f_{sp3}} \times (1 - da), & \text{3 point shot} \\ 2 \times r_{f_{sp}} , & \text{2 point shot foul} \\ 3 \times r_{f_{sp}} , & \text{3 point shot foul} \\ 0, & \text{normal pass} \end{cases} \quad (7)$$

其中,因为 NBA 中每次进攻的平均得分为 1.0765 分,所以当球员  $p$  传球失误(被抢断或传球出界)而失去球权时,  $PESV = -1.0765$ 。当球员  $p$  传球给  $r$ , 球员  $r$  接球后进行 2 分球投篮时,  $PESV = 2 \times r_{f_{sp2}} \times (1 - da)$ 。其中,  $r_{f_{sp2}}$  表示球员  $r$  接到球员  $p$  传球后的 2 分球命中率,而  $da$ (defense affect)表示球员  $r$  投篮时受到的来自防守队员的影响,其中  $da \in [0, 1]$ , 空位投篮时  $da = 0$ , 而投篮被防守队员盖帽时  $da = 1$ 。同样,  $PESV = 3 \times r_{f_{sp3}} \times (1 - da)$  则表示当  $p$  传球给  $r$  后,  $r$  进行 3 分投篮时的传球得分期望。当球员  $r$  投篮被判犯规而罚篮时,传球得分期望值分别为  $2 \times r_{f_{sp}}$ (2 次罚篮)与  $3 \times r_{f_{sp}}$ (3 次罚篮),其中  $r_{f_{sp}}$  表示球员  $r$  的罚球命中率。大多数情况下,当  $p$  传球给  $r$  后,  $r$  并不投篮,而是运球或继续传球,说明球员  $p$  的传球并没有产生得分机会,因此本次传球的得分期望值为 0, 即  $PESV = 0$ 。

首先,传球得分期望值  $PESV$  是一种得分期望计算模型,即球员做出传球后,根据传球后产生的结果来计算这次传球的质量。其次,将式(7)与传球后产生结果的概率结合,可作为传球决策模型(请参见 5.2 节的应用举例)。从式(7)中可以看出,提高  $PESV$  值的原则是尽量传球给投篮命中率高且防守压力较小的球员,即在传球的瞬间取函数  $PESV_{p \rightarrow r}$  的最大值。

### 4.2 球员传球质量的评估

已有的评估球员传球质量的指标是助攻失误比(Assist Turnover Ratio,  $ATR$ ),即助攻数与失误数的比值。但  $ATR$  更侧重于说明球员传球的稳定性(即避免失误的能力),并不能直接反映球员传球的质量。而已有的技术统计中,只有助攻(Assist)这一项技术指标与传球质量相关,但是助攻数并不能全面概括球员的传球质量,因为只有当球员  $p$  传球给  $r$ ,  $r$  投中后,才能算助攻,而当  $r$  投篮不中时,无论  $p$  是多么好的一次传球,也不能算助攻;并且,当  $p$  传球给  $r$ , 而  $r$  投篮犯规后被执行罚篮得分时,理论上的罚篮得分机会与这次传球相关性非常大,但是已有的技术指标或统计都忽略了这次传球的价值。

在式(7)传球得分期望值  $PESV$  的基础上,设一场比赛结束后球员  $p$  一共传球给了  $n$  个队友( $r_1, r_2, r_3, \dots, r_n$ ), 而这  $n$  个队友之间的传球次数分别对应为( $m_1, m_2, m_3, \dots, m_n$ ), 那么若本场比赛球员  $p$  所有传球的  $PESV$  之和为  $S_p$ , 则  $S_p$  的值如式(8)所示:

$$S_p = \sum_{i=1}^n \sum_{j=1}^m (PESV_{p \rightarrow r})_{ij} \quad (8)$$

其中,  $(PESV_{p \rightarrow r})_{ij}$  表示球员  $p$  与第  $i(i \in [1, n])$  名队友之间的第  $j(j \in [1, m_i])$  次传球的 PESV 值。

设球员  $p$  的一次传球的平均传球得分期望值为  $A_p$ , 那么本场比赛的  $A_p$  值由式(9)计算。

$$A_p = \frac{S_p}{\sum_{i=1}^n m_i} = \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} (PESV_{p \rightarrow r})_{ij}}{\sum_{i=1}^n m_i} \quad (9)$$

而在实际应用场景中, 单场比赛的  $A_p$  值并不能准确评估球员的传球质量。假设球员  $p$  一个赛季总共参加了  $t$  场比赛, 每场比赛的平均传球得分期望值分别为  $(A_1, A_2, A_3, \dots, A_t)$ , 设球员整个赛季的平均传球得分期望值为  $AS_p$ , 那么  $AS_p$  值的计算过程由式(10)所示:

$$AS_p = \sum_{k=1}^t A_k = \sum_{k=1}^t \left( \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} (PESV_{p \rightarrow r})_{ij}}{\sum_{i=1}^n m_i} \right)_k \quad (10)$$

相比于传统的助攻数据及助攻失误比数据,  $AS_p$  能够更好地反映球员的传球质量。

### 5 应用场景分析

#### 5.1 传球质量对比赛结果的影响分析

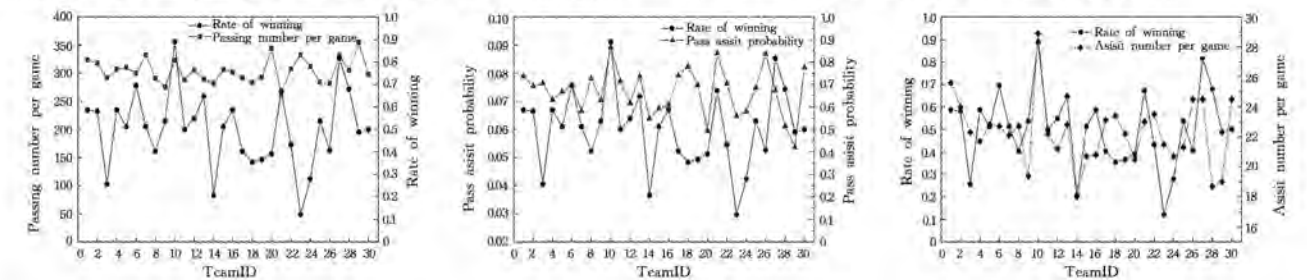
通过数据获取、清洗及统计整理, 本文将 2015—2016 赛季 30 支球队的场均传球数据进行了统计(见表 5), 包括场均传球数、场均助攻数、传球助攻比(产生 1 次助攻需要的传球数量)、传球助攻率(1 次传球转化为助攻的概率)以及球队在 2015—2016 赛季的胜率(赛季总场次除以胜利场次)。通过表 5 数据可统计出: NBA 联盟中 30 支球队的场均助攻数为 22.277, 传球助攻率平均为 7.347%。其中, 场均传球数最多的为 Utah Jazz 队, 达到了场均 354.8 次; 而最少的为 Oklahoma City Thunder 队, 场均传球数量仅为 264 次, 比场均传球数量最多的 Utah Jazz 少了 90.8 次之多。在场均助攻数方

面, Golden State Warriors 达到联盟最高的 28.9 次, 比垫底的 Los Angeles Lakers 的 18 次多出整整 10.9 次。Golden State Warriors 以高达 8.945% 的传球助攻率领跑全联盟, 而 Utah Jazz 的传球助攻率仅为 5.355%。

表 5 2015—2016 赛季各球队场均传球数据汇总

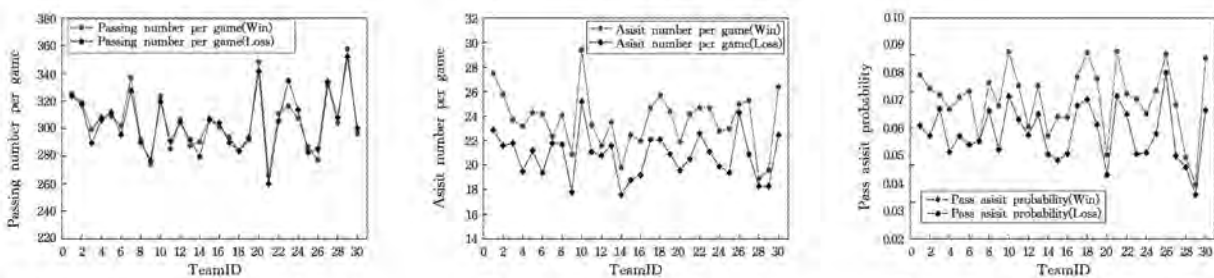
编号	球队名称	场均传球数	场均助攻数	传球助攻比	传球助攻率/%	胜率/%
1	Hawks	324.5	25.6	12.676	7.889	58.537
2	Celtics	318.4	24.0	13.267	7.538	58.025
3	Nets	292	22.3	13.094	7.637	25.610
4	Hornets	307.5	21.7	14.171	7.057	58.537
5	Bulls	310.5	22.8	13.618	7.343	51.220
6	Cavaliers	300.2	22.7	13.225	7.562	69.512
7	Mavericks	332.2	22.1	15.032	6.653	51.220
8	Nuggets	290.8	22.7	12.811	7.806	40.244
9	Pistons	275.5	19.4	14.201	7.042	53.659
10	Warriors	323.1	28.9	11.180	8.945	89.024
11	Rockets	288	22.2	12.973	7.708	50.000
12	Pacers	305.6	21.2	14.415	6.937	54.878
13	Clippers	289.2	22.8	12.684	7.884	64.634
14	Lakers	281.7	18.0	15.650	6.390	20.732
15	Grizzlies	306.9	20.7	14.826	6.745	51.220
16	Heat	302.3	20.8	14.534	6.881	58.537
17	Bucks	291.4	23.1	12.615	7.927	40.244
18	Timberwolves	283.8	23.4	12.128	8.245	35.366
19	Pelicans	293.2	22.2	13.207	7.572	36.585
20	Knicks	344.1	20.5	16.785	5.958	39.024
21	Thunder	264	23.0	11.478	8.712	67.073
22	Magic	307.8	23.5	13.098	7.635	43.210
23	76ers	332.4	21.5	15.460	6.468	12.195
24	Suns	312	20.7	15.072	6.635	28.049
25	Blazers	284.7	21.3	13.366	7.482	53.659
26	Kings	281.9	24.5	11.506	8.691	40.741
27	Spurs	331.7	24.5	13.539	7.386	81.707
28	Raptors	305.4	18.7	16.332	6.123	67.901
29	Jazz	354.8	19.0	18.674	5.355	48.780
30	Wizards	298.2	24.5	12.171	8.216	50.000

本文将球队的胜率与传球数据(包括场均传球数、场均助攻数、传球助攻率)进行对比分析, 其对比结果如图 4 所示。而图 5 中将比赛结果分为输与赢两类进行分析。



(a) The relationship between the passing number per game and the winning probability (b) The relationship between the pass assist probability and the winning probability (c) The relationship between the assist number per game and the winning probability

图 4 球队胜率与传球数据之间的关系



(a) The comparison of the pass number per game while win and loss (b) The comparison of the assist number per game while win and loss (c) The comparison of pass assist probability per game while win and loss

图 5 2015—2016 赛季传球数据对比赛结果(赢或输)的影响分析

综合观察图 4 和图 5 可以发现,场均传球次数与比赛结果之间并无直接关系(或相关性很小),但是传球质量(传球助攻转化率与场均助攻数越高,传球质量越高)与比赛结果的相关性较大,即传球质量越高,赢球的可能性越大。图 4 和图 5 从侧面表明了研究传球质量的重要性。

### 5.2 基于 PESV 值的传球路线选择

图 6 给出了 2015-2016 赛季总决赛时两支队伍比赛的真实场景示例,此时常规赛 MVP Stephen Curry 持球,并且遭到 Kyrie Irving 与 J R Smith 两位球员的包夹;此时 Stephen Curry 的传球路线一共有 4 条,那么选择哪条传球路线的传球得分期望值 PESV 最高? 结合式(7),计算 PESV 值时需要知道接球球员的投篮命中率( $r_{jgp2}, r_{jgp3}$ )及所受到的防守压力。结合第 3 节的传球网络构建及可视化,可从勇士队的传球网络中以 Stephen Curry 为顶点抽取传球网络子图,图 7 所示是以库里为顶点的传球网络子图。以 Stephen Curry 为出发顶点,可以得到与其他所有队员之间的传球相关信息,包括对应球员的投篮命中率。此时,只需根据当时的防守参数  $da$  的值,便能计算出 4 条传球路线的 PESV 值。

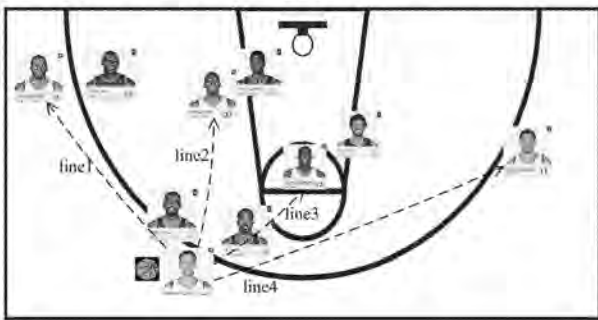


图 6 库里应该选择哪条传球路线?

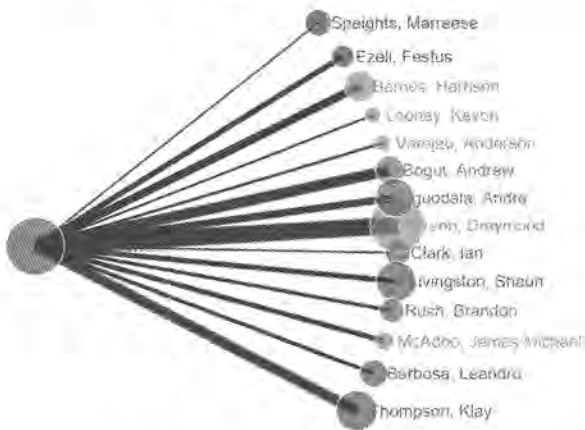


图 7 以库里为顶点的传球网络子图

当 Stephen Curry 选择传球路线 1 时(即传球给队友 Andre Iguodala),传球路线需要经过 Kyrie Irving 与 LeBron James,传球有被抢断的风险,并且队友 Andre Iguodala 受到 LeBron James 的防守压力较大,没有投篮的空间。假设选择传球路线 1 时被抢断的概率为 0.15,那么选择传球路线 1 的 PESV 期望值为:

$$PESV_{line1}^{expected} = -1.0765 \times 0.15 + 0 \times (1 - 0.85) = -0.161476$$

当选择传球路线 2 时(即传球给队友 Festu Ezeli),传球

路线需经过 Kyrie Irving 与 J R Smith 的包夹,传球被抢断的风险较高,并且即便传球成功,Festu Ezeli 投篮面对的防守压力也很大。假设选择传球路线 2,被抢断的概率为 0.25,队友所受防守压力值为 0.6,投篮犯规的概率为 0.3,那么选择传球路线 2 的 PESV 期望值为:

$$PESV_{line2}^{expected} = -1.0765 \times 0.25 + 2 \times 0.478 \times (1 - 0.6) \times (1 - 0.25) + 2 \times 0.53 \times 0.3 \times (1 - 0.25) = -0.269125 + 0.2868 + 0.2385 = 0.256175$$

其中,0.478 为 Festu Ezeli 接到 Stephen Curry 后的投篮命中率,0.53 为 Festu Ezeli 的罚球命中率。

当选择传球路线 3 时(即传球给队友 Draymond Green),传球路线需经过防守人 J R Smith,被抢断的概率为 0.2; Draymond Green 投篮时受到的防守压力值为 0.3(防守人 Kevin Love 的防守能力较弱),投篮犯规的概率为 0.2。此时,选择传球路线 3 的 PESV 期望值为:

$$PESV_{line3}^{expected} = -1.0765 \times 0.2 + 2 \times 0.521 \times (1 - 0.3) \times (1 - 0.2) + 2 \times 0.696 \times 0.3 \times (1 - 0.2) = -0.2153 + 0.58352 + 0.33408 = 0.7023$$

其中,0.521 为 Draymond Green 的 2 分球投篮命中率,0.696 为罚球命中率。

当选择传球路线 4 时(即传球给队友 Klay Thompson),传球路线上没有防守队友,被抢断的概率趋近于 0,并且 Klay Thompson 此时处于空位状态,受到的防守压力也趋近于 0,当 Stephen Curry 传球给 Klay Thompson 时的 3 分命中率为 0.455,那么选择传球路线 4 时的 PESV 期望值为:

$$PESV_{line4}^{expected} = 3 \times 0.455 = 1.365$$

因此,基于以上论述,Stephen Curry 最佳的传球路线选择为 line4,line3 次之,line1 最差。以上是对 Stephen Curry 传球前的 PESV 期望值的分析。如果 Stephen Curry 最终的传球选择是 line2,传球没有被抢断,并且 Festu Ezeli 的投篮造成了防守队员的犯规,此时的 PESV 的真实值为:

$$PESV_{line2} = 2 \times 0.53 = 1.06$$

传球路线的 PESV 期望值虽然不能在实际比赛中帮助球员选择出具体的传球路线(因为在真实的比赛中球员的每次传球只能根据自己的瞬时判断,近乎本能地做出传球路线的选择),但是在如 NBA 2K16 等篮球游戏中,以上传球路线中的 PESV 计算方法则有很好的应用价值,能够帮助提升游戏中人物传球的智能程度。

### 5.3 球员平均传球得分期望值的计算

以华人球员林书豪 2015-2016 赛季常规赛期间传球数据为例进行分析,其传球数据如表 6 所列。

2015-2016 赛季林书豪一共传球 2479 次,共产生 234 次助攻,传球助攻率为 9.44%。2479 次传球中,共产生 470 次 2 分球投篮机会,275 次 3 分球投篮机会。那么,林书豪 2479 次传球中产生投篮机会次数为 745,即传球投篮机会产生率为 30%。结合式(7)与表 6 数据,林书豪除掉失误后的传球得分期望值总和为:

$$PESV_{sum} = \sum 2 \times r_{jgp2} \times (1 - da) + \sum 3 \times r_{jgp3} \times (1 - da) + (-1.0765 \times turnover\_num)$$

表6 林书豪 2015—2016 赛季传球数据汇总

接球人	传球次数	助攻次数	两分球投篮数	两分球命中率/%	三分球投篮数	三分球命中率/%
Kaminsky	397	35	58	51.7	39	35.9
Walker	394	23	65	53.8	51	35.3
Batum	316	16	38	55.3	38	26.3
Hawes	316	32	63	42.9	30	43.3
Lamb	258	31	62	58.1	39	38.5
Williams	239	26	31	38.7	37	43.2
Jefferson	169	25	71	50.7	0	0
Zeller	127	17	42	40.5	2	0
Daniels	69	7	4	25	12	58.3
Lee	65	4	10	40	7	14.3
Hairston	51	9	7	57.1	15	40
Roberts	37	2	7	42.9	4	25
Hansbrough	26	5	8	62.5	0	0
Kidd-Gilchrist	13	2	4	50	0	0
Harrison	2	0	0	0	1	0

由于在赛后的总体统计数据中无法对每次传球投篮时的防守压力参数  $da$  进行确认,在进行球员对比时,为了简化,可将参数  $da$  统一设置为 0。由于林书豪整个赛季中失误 150 次,因此可计算林书豪整个赛季中所有传球的  $PESV$  之和为:

$$PESV_{sum} = 466 + 303 + (-1.0765 \times 150) = 607.525$$

由于林书豪在 2015—2016 赛季总共出场 78 次,因此林书豪的场均  $PESV$  值为 7.78,平均每次传球的  $PESV$  值为 0.245。

**结束语** 大数据在竞技体育方面的相关应用研究还处于探索阶段。本文发现篮球比赛中除得分、篮板、助攻、抢断、盖帽、失误、犯规、投篮命中率、出场时间等常规技术统计外,还缺乏对传球数据的记录,更缺乏对传球质量量化方面的研究。首先,本文将球员之间的传球关系进行关联,发现随着传球数据量的不断增加,最终形成了一张稳定的传球网络图。其次,在对传球网络进行可视化的基础上,研究了传球质量的量化方法,并以此对球员的传球质量进行评估。最后,列举了传球质量模型的几种应用场景,包括传球质量对比赛结果的影响分析、基于  $PESV$  值的传球路线选择,以及以华人球员林书豪为例来计算球员整个赛季的传球得分期望值。相比于传统的助攻数与失误数的比值  $ATR$ (助攻失误比),本文提出的传球质量评估方法  $PESV$  能更直接地反映球员传球的质量。

下一步工作主要集中在以下 3 个方面:

(1) 高效的获取传球数据的方法的研究。由于已有的传球数据只能通过人工现场统计或录像分析的方法得到,效率低下且不能避免人为错误。因此,需要与图像识别及视频处理等技术相结合,研究自动从比赛录像中提取传球数据的方法。

(2) 本地化的基于传球网络的应用开发。中国的 CBA 甚至是中国男篮,对常规技术统计数据的价值挖掘及应用都停留在初级阶段。因此,研究基于传球网络的应用软件,从而简化传球数据的获取到数据价值转化的过程,是推动篮球竞技数据化、智能化的关键。

(3) 将传球数据质量与其他技术统计进行关联分析,这要求在已有的传球数据中扩展传球类型、速度、失误概率、传球距离等信息,并挖掘传球数据与球员位置、得分、失误、助攻等其他技术统计数据与传球质量之间的关联。

## 参考文献

- [1] GHEMAWAT S, GOBIOFF H, LEUNG S T. The google file system[C]// Proceedings of 19th ACM Symposium on Operating System Principles. New York: ACM, 2003: 29-43.
- [2] DEAN J, GHEMAWAT S. MapReduce: Simplified data processing on large clusters[C]// Proceedings of the Conference on Operating System Design and Implementation (OSDI). New York: ACM, 2004: 137-150.
- [3] ZAHARIA M, CHOWDHURY M, FRANKLIN M J, et al. Spark: cluster computing with working sets[C]// Usenix Conference on Hot Topics in Cloud Computing. USENIX Association, 2010: 1765-1773.
- [4] ZAHARIA M, CHOWDHURY M, DAS T, et al. Resilient distributed datasets: a fault-tolerant abstraction for in-memory cluster computing[C]// Usenix Conference on Networked Systems Design and Implementation. 2012: 141-146.
- [5] XIN R S, GONZALEZ J E, FRANKLIN M J, et al. GraphX: a resilient distributed graph system on Spark[C]// International Workshop on Graph Data Management Experiences and Systems. 2013: 1-6.
- [6] LOW Y, BICKSON D, GONZALEZ J, et al. Distributed GraphLab: a framework for machine learning and data mining in the cloud[J]. Proceedings of the VLDB Endowment, 2012, 5(8): 716-727.
- [7] HAN M, DAUDJEE K. Giraph unchained: barrierless asynchronous parallel execution in pregel-like graph processing systems[J]. Proceedings of the VLDB Endowment, 2015, 8(9): 950-961.
- [8] YAN Y L, DONG Y H, HE X M, et al. FSMBUS: A Frequent Subgraph Mining Algorithm in Single Large-Scale Graph Using Spark[J]. Journal of Computer Research and Development, 2015, 52(8): 1768-1783. (in Chinese)  
严玉良,董一鸿,何贤芒,等. FSMBUS: 一种基于 Spark 的大规模频繁子图挖掘算法[J]. 计算机研究与发展, 2015, 52(8): 1768-1783.
- [9] CERVONE D, D'AMOUR A, BORN L, et al. POINTWISE: Predicting points and valuing decisions in real time with nba optical tracking data[C]// Proceedings of the 8th MIT Sloan Sports Analytics Conference. Boston, MA, USA, 2014: 1-9.
- [10] MAHESWARAN R, CHANG Y H, SU J, et al. The three dimensions of rebounding[C]// Proceedings of the 8th MIT Sloan Sports Analytics Conference. Boston, MA, USA, 2014: 1-7.
- [11] MAYMIN P. Acceleration in the NBA: Towards an algorithmic taxonomy of basketball plays[C]// Proceedings of the 7th MIT Sloan Sports Analytics Conference. Boston, MA, USA, 2013: 1-7.
- [12] GOLDMAN M, RAO J M. Live by the Three, Die by the Three? The Price of Risk in the NBA[C]// Proceedings of the 7th MIT Sloan Sports Analytics Conference. Boston, MA, USA, 2013: 1-15.
- [13] FRANKS A, MILLER A, BORN L, et al. Counterpoints:

(下转第 201 页)

低当前算法的复杂度,尝试将自然邻居思想应用于图像处理、大数据等热门领域,将自然邻居思想发展得更为成熟。

### 参 考 文 献

- [1] AMBERT K H, COHEN A M. k-Information gain scaled nearest neighbors: a novel approach to classifying protein-protein interaction-related documents[J]. *IEEE/ACM Transactions on Computational Biology & Bioinformatics*, 2012, 9(1): 305-310.
- [2] BRIAN M, CAROLINA G, GERT L. Contextual Object Localization With Multiple Kernel Nearest Neighbor [J]. *IEEE Transactions on Image Processing*, 2011, 20(2): 570-585.
- [3] SALVADOR G, JOAQUIN D, JOSE R C. Prototype Selection for Nearest Neighbor Classification: Taxonomy and Empirical Study [J]. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2012, 34(3): 417-435.
- [4] STEVENS S S. Mathematics, Measurement, and Psychophysics [M]//*Handbook of Experimental Psychology*. 1951: 1-49.
- [5] KORN F, MUTHUKRISHNAN S. Influence Sets Based on Reverse Nearest Neighbor Queries [J]. *ACM SIGMOD Record*, 2000, 29(2): 201-212.
- [6] WANG J, NESKOVIC P, COOPER L. Improving Nearest Neighbor rule with a simple adaptive distance measure [J]. *Pattern Recognition Letter*, 2007, 28(2): 207-213.
- [7] SALVADOR G, JOAQUIN D, JOSE R C. Prototype Selection for Nearest Neighbor Classification: Taxonomy and Empirical Study [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(3): 417-435.
- [8] GHOSH A K. On Optimum Choice of k in Nearest Neighbor classification [J]. *Computational Statistics & Data Analysis*, 2006, 50(11): 3113-3123.
- [9] HASTIE T, TIBSHIRANI R. Discriminant Adaptive Nearest neighbor Classification [J]. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 1996, 18(6): 607-616.
- [10] GHOSH A K, ANIL K. On Nearest Neighbor Classification sing Adaptive Choice of k [J]. *Computational & Graphical Statistics*, 2007, 16(2): 482-502.
- [11] DOMENICONI C, PENG J, GUNOPULOS D. Locally Adaptive metric Nearest-Neighbor Classification [J]. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 2002, 24(9): 1281-1285.
- [12] BHATTACHARYA G, GHOSH K, CHOWDHURY A S. An affinity-based new local distance function and similarity measure for kNN algorithm [J]. *Pattern Recognition Letters*, 2012, 33(3): 356-363.
- [13] YIU M L, MAMOULIS N. Reverse Nearest Neighbors Search in Ad Hoc Subspaces [J]. *IEEE Trans. on Knowledge and Data Engineering*, 2007, 19(3): 412-426.
- [14] WANG S S, CHAI S, LV Q N. A Pruning Based Continuous RkNN Query Algorithm for Large k [J]. *Chinese Journal of Electronics*, 2012, 21(3): 523-552.
- [15] ZHANG Y. Study on Classification algorithm based on natural nearest neighbor [D]. Chongqing: Chongqing University, 2015. (in Chinese)  
张莹. 基于自然最近邻居的分类算法研究[D]. 重庆: 重庆大学, 2015.
- [16] HUANG J L. Study on non-parametric clustering based on natural nearest neighborhood [D]. Chongqing: Chongqing University, 2014. (in Chinese)  
黄金龙. 基于自然最近邻的无参聚类算法研究[D]. 重庆: 重庆大学, 2014.
- [17] TANG H. An outlier detection algorithm based on natural nearest neighbor [D]. Chongqing: Chongqing University, 2014. (in Chinese)  
唐汇. 基于自然最近邻居的离群检测算法研究[D]. 重庆: 重庆大学, 2014.
- [18] INKAYA T, KAYALIGIL S, ÖZDEMIREL N E. An adaptive neighbourhood construction algorithm based on density and connectivity [J]. *Pattern Recognition Letters*, 2015, 52: 17-24.
- (上接第 182 页)
- Advanced defensive metrics for nba basketball[C]//*Proceedings of the 9th MIT Sloan Sports Analytics Conference*. Boston, MA, USA, 2015: 1-8.
- [14] WIENS J, GUHA BALAKRISHNAN J B, GUTTAG J. To Crash or Not To Crash: A quantitative look at the relationship between offensive rebounding and transition defense in the NBA [C]//*Proceedings of the 7th MIT Sloan Sports Analytics Conference*. Boston, MA, USA, 2013: 1-7.
- [15] REN L, DU Y, MA S, et al. Visual analytics towards big data [J]. *Journal of Software*, 2014, 25(9): 1909-1936. (in Chinese).  
任磊, 杜一, 马帅, 等. 大数据可视分析综述[J]. *软件学报*, 2014, 25(9): 1909-1936.
- [16] HERMAN I, MELANCON G, MARSHALL M S. Graph visualization and navigation in information visualization: A survey [J]. *IEEE Transaction on Visualization and Computer Graphics*, 2000, 6(1): 24-43.
- [17] ZHANG X, YUAN X R. Treemap visualization [J]. *Journal of Computer-Aided Design & Computer Graphics*, 2012, 24(9): 1113-1124.
- [18] BALZER M, DEUSSEN O. Voronoi Treemaps [C]//*IEEE Symposium on Information Visualization*. Los Alamitos: IEEE, 2005: 49-56.
- [19] GOU L, ZHANG X. Treemapviz: Revealing patterns of networks over tree structures [J]. *IEEE Transaction on Visualization and Computer Graphics*, 2011, 17(12): 2449-2458.
- [20] ZHANG T, YU J, LIAO B, et al. The Construction and Analysis of Pass Network Graph Based on GraphX [J]. *Journal of Computer Research and Development*, 2016, 53(12): 2729-2752. (in Chinese)  
张陶, 于炯, 廖彬, 等. 基于 GraphX 的传球网络构建及分析研究 [J]. *计算机研究与发展*, 2016, 53(12): 2729-2752.