基于模糊逻辑的 NBA 球员匹配专家系统

万瑞 19307130250

2022.6.4

目录

摘要

通过用户描述球员的多种表现要求，本专家系统将推荐出较为匹配的球员名单。在实现方面，采用了基于模糊推理的专家系统，输入每个球员的过往数据，评估球员的表现，再计算描述球员和过往球员间的关联度，最后依据关联度给出匹配名单。

关键词　专家系统、模糊推断、NBA

1 介绍

National Basketball Association (NBA) 是指北美职业篮球联盟。每一年，联盟将通过选秀的方式招纳新秀，每位球员参加选秀时，会被赋予对应的模版，参照过往球员，给出职业路径的窥见。本专家系统即可以通过输入新秀的身体特征和期望的使用环境，来匹配相似度较高的过往球员

本专家系统的架构设计如图1下:

- 用户提供球员信息，球员数据参数即扮演位置，球员使用环境包括防守型、投篮型、篮板型、控场型、突破型。过往球员数据通过模糊逻辑得出使用环境，然后计算用户球员使用环境和过往球员使用环境的匹配程度，给出推荐排名。
- 本系统的专家和开发人员都由我承担，通过分析球员数据库，给出模糊推理逻辑，并录入专家系统知识库和规则库中，推理引擎选择了开源 Fuzzy Logic Controller (FLC) 项目 jFuzzyLogic[1, 2]，使用 Fuzzy Control Language (FCL) 来构建模糊规则 [3]

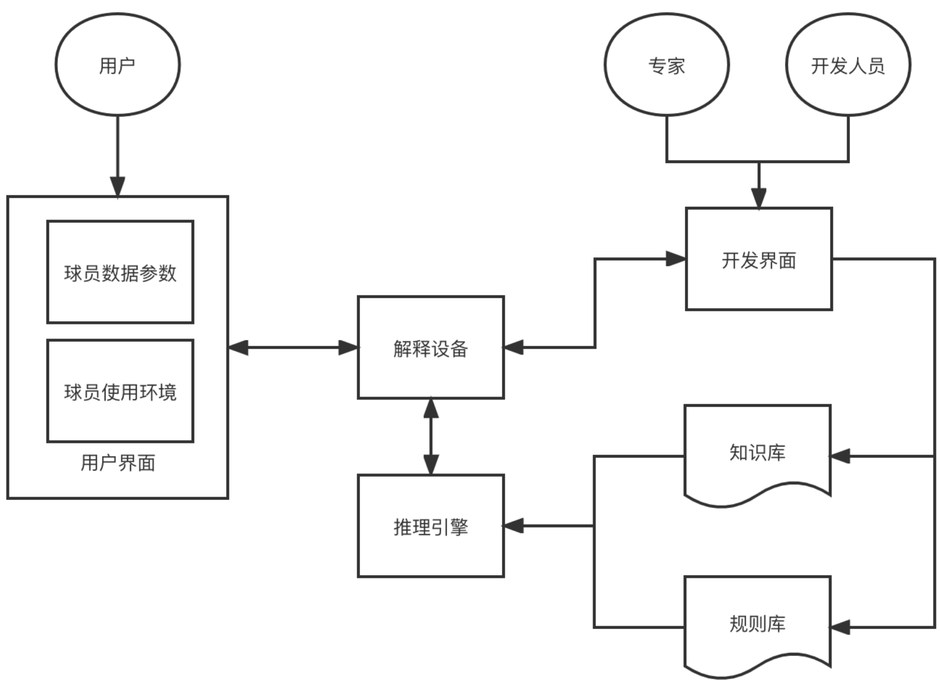


图 1: structure

2 具体实现

2.1 数据来源

过往球员数据来自于 kaggle 平台的 2021-2022 NBA Player Stats - Playoffs 数据集 [4]，共 715 条数据，30 种属性，记录了球员的身体特征，以及 2021-2022 季后赛场均表现。提取单条数据展示如下；

```json
{
    "Rk": 77,
    "Player": "Aaron Gordon",
    "Pos": "PF",
    "Age": 26,
    "Tm": "DEN",
    "G": 5,
    "GS": 5,
    "MP": 32.0,
    "FG": 4.6,
    "FGA": 10.8,
    "FG%": 0.426,
    "3P": 0.6,
    "3PA": 3.0,
    "3P%": 0.2,
    "2P": 4.0,
    "2PA": 7.8,
    "2P%": 0.513,
    "eFG%": 0.454,
    "FT": 4.0,
    "FTA": 5.6,
    "FT%": 0.714,
    "ORB": 3.6,
    "DRB": 3.6,
    "TRB": 7.2,
    "AST": 2.6,
    "STL": 0.4,
    "BLK": 1.2,
    "TOV": 1.6,
    "PF": 2.8,
    "PTS": 13.8
}
```

Listing 1: Aaron Gordon.json

2.2 问题设计

1. 您希望球员扮演的角色（多选）

- point guard
- shooting guard
- small forward
- power forward
- center

2. 您期望球员的投射能力 (拖动条)
3. 您期望球员的组织能力 (拖动条)
4. 您期望球员的防守能力 (拖动条)
5. 您期望球员的攻框能力 (拖动条)
6. 您期望球员的篮板能力 (拖动条)

2.3 系统实现设计

本系统选择模糊专家系统的原因是，球员的使用环境并不是单纯的二值布尔逻辑，而是多种使用环境的多值逻辑。并且球员的自身数据也不是单纯的二值布尔逻辑，真实情况是反映了该球员在整个联盟中的水平高低。比如一位球员的抢断数据 (Steals per game, STL) 为 0.5，则意味着在联盟中为中等水平，经过模糊逻辑中可以对防守型、控场型有少量提升作用。所以本专家系统中采用了模糊逻辑对每个球员进行使用环境分析，然后给出各环境的适配度。

然后将计算出的 n 维环境向量和用户定义的环境向量，通过欧式距离计算相似度，即可得出匹配度排名。

整个流程如图2所示。

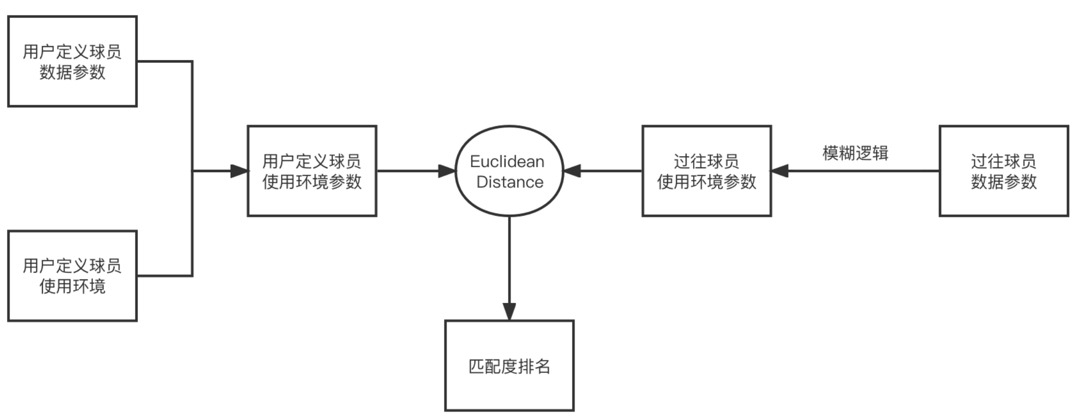


图 2: workflow

2.4 定义语言变量以及确定模糊集

2.4.1 输入变量

系统中我们有 24 个语言变量，如附录表2所示。然后需要确定模糊集，我们的方式是参考变量的统计信息，如附录图5所示。这里以 MP 为例划分为：

- very low(VL) : 0 - 4.6
- low(L) : 0 -13.54
- medium(M) : 11.76 - 26.68
- high(H) : 24.46 - 44.00
- very high(VH) : 33.78 - 44.00

翻译为 jFuzzyLogic 所使用的 FCL 语言即为

```
FUZZIFY MP
  TERM VL := (0.0, 1) (2.96, 1) (4.600000000000001, 0);
  TERM L := (0.0, 0) (3.8, 1) (10.260000000000003, 1) (13.540000000000001, 0);
  TERM M := (11.76, 0) (15.500000000000002, 1) (21.580000000000002, 1)
(26.680000000000003, 0);
  TERM H := (24.46, 0) (29.02, 1) (35.7, 1) (44.0, 0);
  TERM VH := (33.78000000000001, 0) (37.24, 1) (44.0, 1);
END_FUZZIFY
```

Listing 2: MP scope

使用 jFuzzyLogic 工具即可获得模糊集，如图3所示

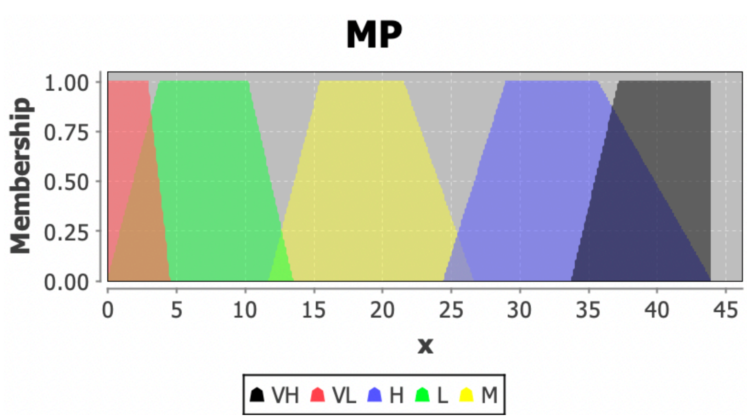


图 3: fuzzy set of MP

全部语言变量及值范围参见附录表3

2.4.2 输出变量

系统中有 5 个输出变量，其定义和值范围描述如表1所示

变量名称	变量简称	语言值	符号	值范围
防守型	defensive	very bad	VB	[0, 0.35]
投篮型	shooting	bad	B	[0.15, 0.45]
篮板型	rebounding	medium	M	[0.3, 0.7]
控场型	leading	good	G	[0.55, 0.85]
突破型	breaking	very good	VG	[0.65, 1]

表 1: output variable

以 defensive 为例的 jFuzzyLogic 语言如下

```
DEFUZZIFY defensive
  TERM VB := (0, 1) (0.2, 1) (0.35, 0);
  TERM B := (0.15, 0) (0.3, 1) (0.45, 0);
  TERM M := (0.3, 0) (0.5, 1) (0.7, 0);
  TERM G := (0.55, 0) (0.7, 1) (0.85, 0);
  TERM VG := (0.65, 0) (0.8, 1) (1,1);
  METHOD : COG;
  DEFAULT := 0.5;
END_DEFUZZIFY
```

Listing 3: defensive scope

构建得到的模糊集如图4所示

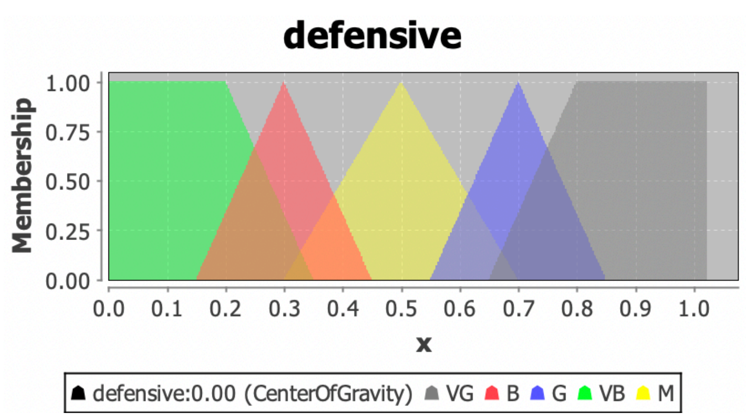


图 4: fuzzy set of defensive

2.5 构造模糊规则

在构建模糊规则时，参考了篮球知识网站 [5, 6]，以制定更专业的模糊规则

以 leading 为例，输入参数中 AST、TOV、STL、ORB、DRB 对其具有较大的参考价值，比如

- 助攻数与控场能力呈正相关
- 失误数与控场能力呈负相关
- 当抢断数较高时，控场能力一般比较好
- 当进攻篮板数较高时，控场能力一般比较好
- 当防守篮板数较高时，控场能力一般比较好

在构建为 jFuzzyLogic 的 fcl 语言后如下

```
    RULEBLOCK leading
      AND : MIN;  // Use 'min' for 'and'
      ACT : MIN;  // Use 'min' activation method
      ACCU : MAX; // Use 'max' accumulation method

      RULE 1 : IF AST IS VL THEN leading IS VB;
        RULE 2 : IF AST IS L THEN leading IS B;
        RULE 3 : IF AST IS M THEN leading IS M;
        RULE 4 : IF AST IS H THEN leading IS G;
        RULE 5 : IF AST IS VH THEN leading IS VG;

        RULE 6 : IF TOV IS VL THEN leading IS VG;
        RULE 7 : IF TOV IS L THEN leading IS G;
        RULE 8 : IF TOV IS M THEN leading IS M;
        RULE 9 : IF TOV IS H THEN leading IS B;
        RULE 10 : IF TOV IS VH THEN leading IS VB;

        RULE 11 : IF STL IS VH OR STL IS H THEN leading IS G;
        RULE 12 : IF ORB IS VH OR ORB IS H THEN leading IS G;
        RULE 13 : IF DRB IS VH THEN leading IS G;
    END_RULEBLOCK
```

Listing 4: rule of leading

其余模糊规则可在附录6.3查看，共 69 条规则。

3 相似度比较

在我们的专家系统中，对每个球员的 24 个输入参数，评估出了 5 个输出参数，结果实例如下

```
"assessment":{
    "breaking":0.8125295761527721,
    "defensive":0.8125295761527721,
```

```
    "leading":0.49999999999999983,
    "rebounding":0.8125295761527721,
    "shooting":0.49999999999999956
  },
```

Listing 5: assessment of Aaron Gordon

然后根据用户输入的球员参数，作欧式距离计算 *user_player* 为用户输入的五个角度的参数值，*src_player* 为每个球员的 assessment 值从 [0-1] 标准化到 [0-100]

$$distance = \sqrt{\sum_{i=1}^{n} (user\_player_i - src\_player_i)^2}.$$

这样对每个球员进行评分排名，选取 distance 结果最小的三个球员，作为最相似的球员。

4 模拟实验

系统的可视化选用了 Flask 搭建的网页。

输入界面如附录图6所示，会传给后端如下信息：

```
{
        "pos":[PG, SF],
    "breaking":68,
    "defensive":50,
    "leading":41,
    "rebounding":57,
    "shooting":62
  }
```

Listing 6: user input

输出结果如附录图7所示

5 总结

本专家系统还有如下改进方向

- 本专家系统由于输入参数有 20 种，在规划模糊规则时，需要对应的条数较多，可以进一步再精细化输出结果的变量范围。
- 其次输出参数存在部分关联性，比如在本专家系统的模糊规中的防守型和篮板型，但在实际情况中，防守型可能更指赛场上全局的影响，而篮板型的应用会在赛局末端强调，实际上会有时间和空间上的区别。如果专家系统能够从时间和球员位置上进行进一步精细化，结果会更加可靠。

- 可以分析和展现对结果影响较大的输入参数，会更好的说明理由，但由于 jFuzzyLogic 是一个封装的工具，可能需要另外寻求方法，从外部分析计算过程。

总的来说，本专家系统操作简单，正确响应了用户的输入，通过几个问题，解决了用户的需求，在推荐匹配球星上具有一定的参考价值。

6 附录

6.1 部分图片

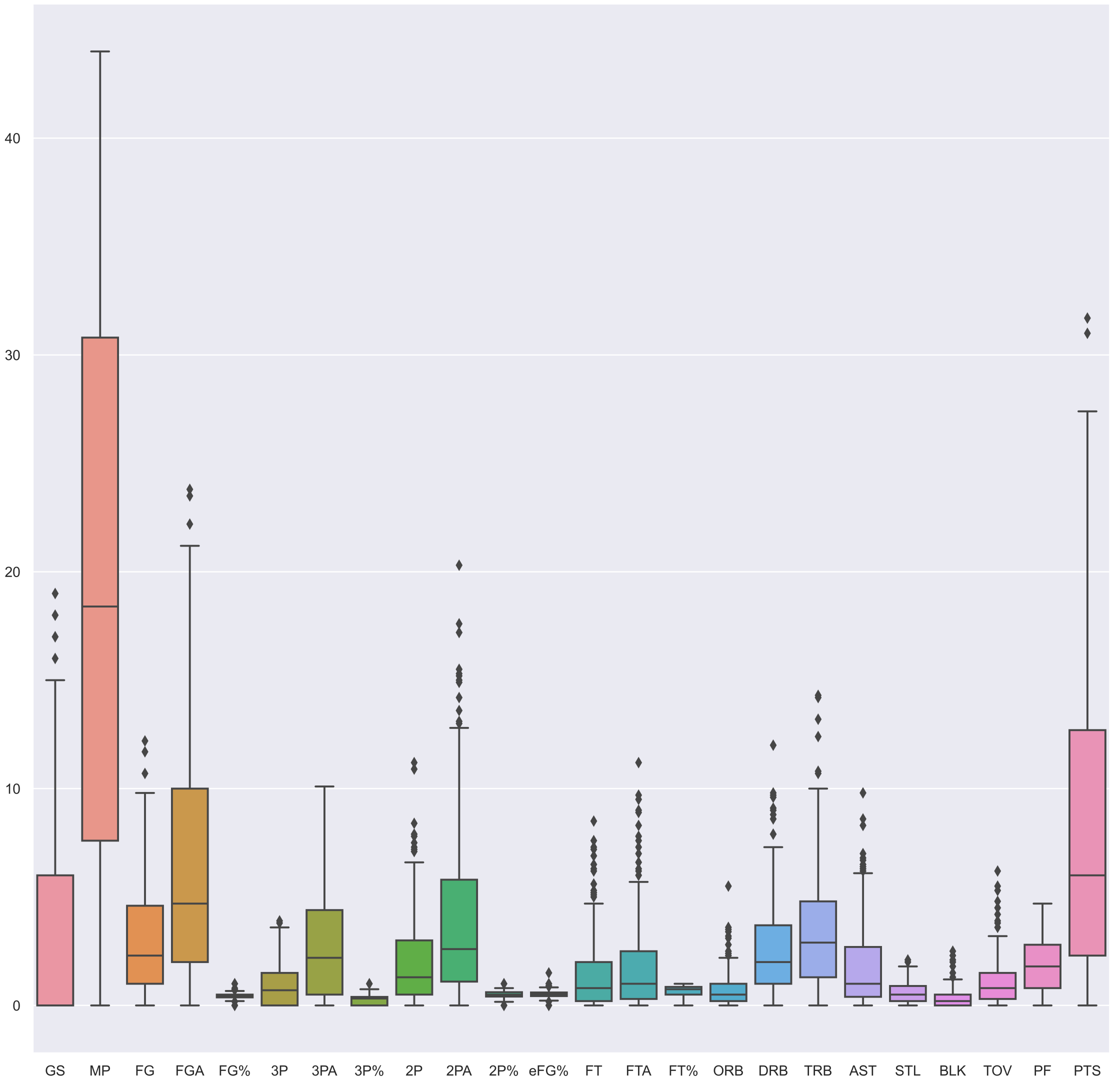


图 5: variable statistics

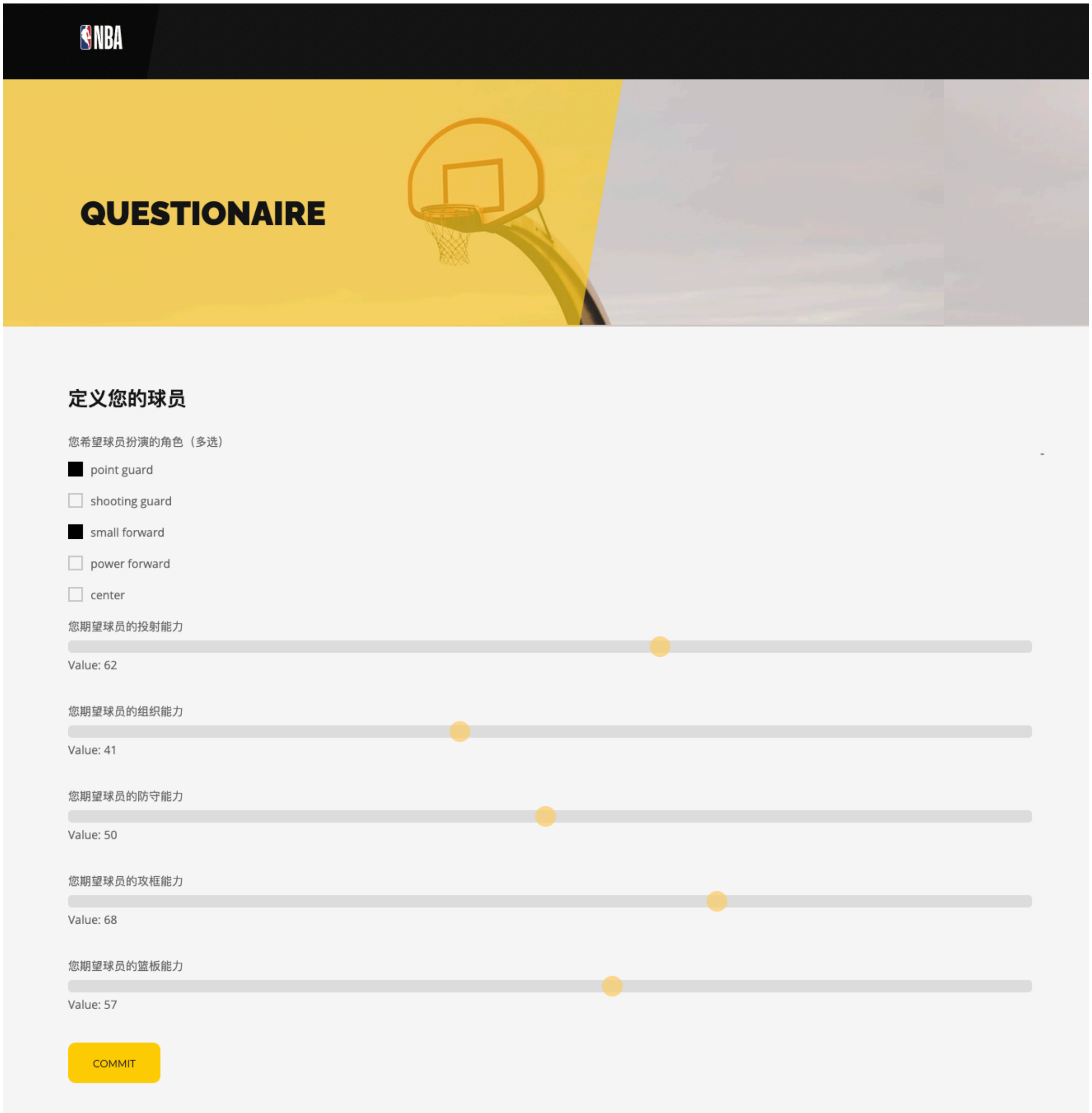

NBA
QUESTIONAIRE
定义您的球员
您希望球员扮演的角色（多选）
point guard
shooting guard
small forward
power forward
center
您期望球员的投射能力
Value: 62
您期望球员的组织能力
Value: 41
您期望球员的防守能力
Value: 50
您期望球员的攻框能力
Value: 68
您期望球员的篮板能力
Value: 57
COMMIT

图 6: 网站输入

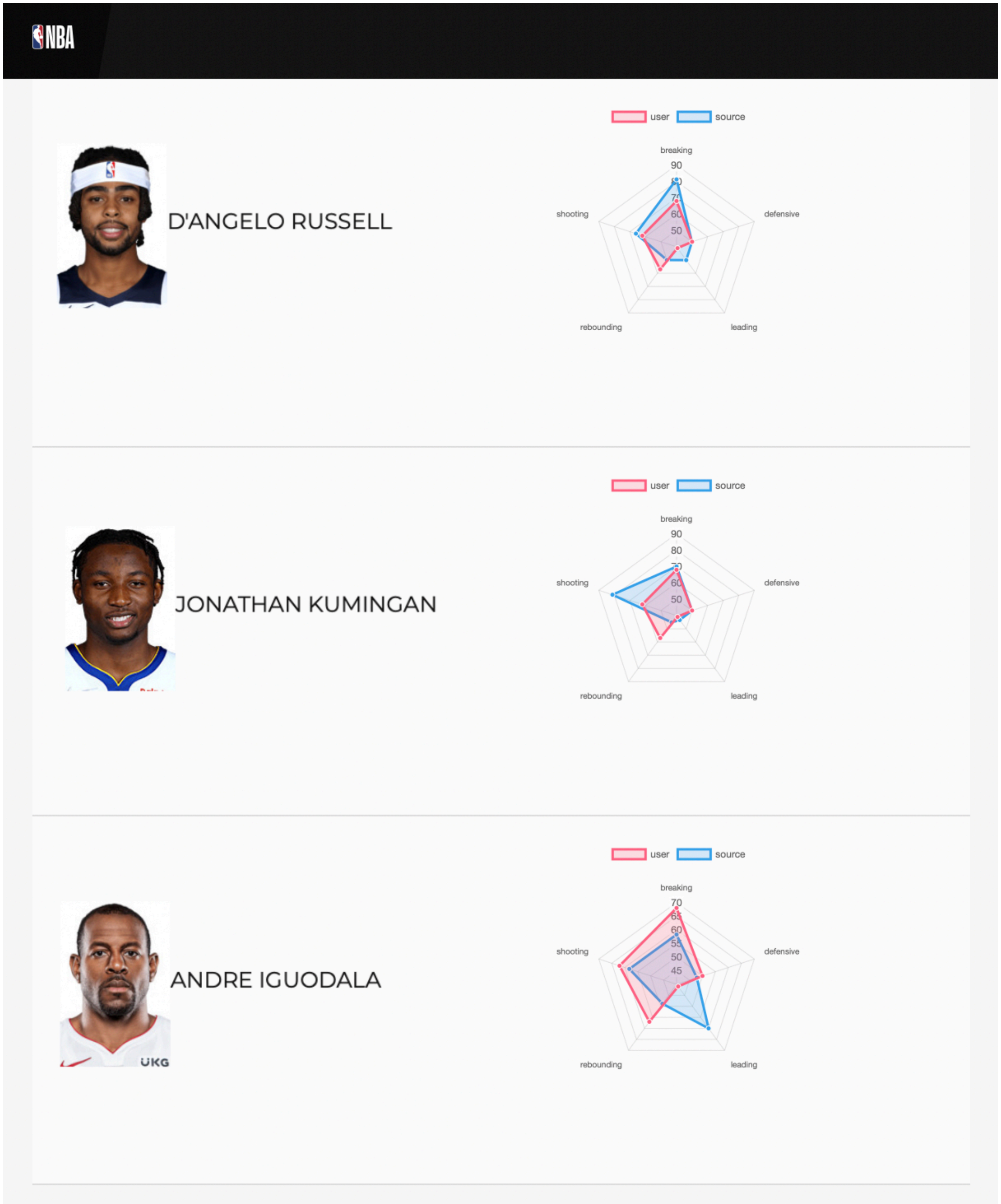

NBA
D'ANGELO RUSSELL
user
source
breaking
shooting
defensive
rebounding
leading
JONATHAN KUMINGAN
ANDRE IGUODALA

图 7: 网站输出

6.2 部分表格

变量名称	变量简称	变量名称	变量简称
Games started	GS	Free throws per game	FT
Minutes played per game	MP	Free throw attempts per game	FTA
Field goals per game	FG	Free throw percentage	FT%
Field goal attempts per game	FGA	Offensive rebounds per game	ORB
Field goal percentage	FG%	Defensive rebounds per game	DRB
3-point field goals per game	3P	Total rebounds per game	TRB
3-point field goal attempts per game	3PA	Assists per game	AST
3-point field goal percentage	3P%	Steals per game	STL
2-point field goals per game	2P	Blocks per game	BLK
2-point field goal attempts per game	2PA	Turnovers per game	TOV
2-point field goal percentage	2P%	Personal fouls per game	PF
Effective field goal percentage	eFG%	Points per game	PTS

表 2: input variable

Begin of Input Variable Scope

GS		
语言值	符号	值范围
very low	VL	[0.0, 0.0]
low	L	[0.0, 0.0]
medium	M	[0.0, 4.0]
high	H	[3.0, 19.0]
very high	VH	[6.0, 19.0]
MP		
语言值	符号	值范围
very low	VL	[0.0, 4.600000000000001]
low	L	[0.0, 13.540000000000001]
medium	M	[11.76, 26.680000000000003]
high	H	[24.46, 44.0]
very high	VH	[33.78000000000001, 44.0]
FG		
语言值	符号	值范围
very low	VL	[0.0, 0.7]

Continuation of Input Variable Scope

low	L	[0.0, 1.6]
medium	M	[1.4, 3.5]
high	H	[3.0600000000000023, 12.2]
very high	VH	[5.1800000000000015, 12.2]
FGA		
语言值	符号	值范围
very low	VL	[0.0, 1.7]
low	L	[0.0, 3.4]
medium	M	[3.0, 7.680000000000001]
high	H	[6.920000000000004, 23.8]
very high	VH	[12.0, 23.8]
FG%		
语言值	符号	值范围
very low	VL	[0.0, 0.333]
low	L	[0.0, 0.4114]
medium	M	[0.4, 0.484]
high	H	[0.4766, 1.0]
very high	VH	[0.5418000000000002, 1.0]
3P		
语言值	符号	值范围
very low	VL	[0.0, 0.0]
low	L	[0.0, 0.3]
medium	M	[0.25999999999999945, 1.2]
high	H	[1.0, 3.9]
very high	VH	[1.8, 3.9]
3PA		
语言值	符号	值范围
very low	VL	[0.0, 0.3]
low	L	[0.0, 1.1]
medium	M	[0.9599999999999994, 3.6]
high	H	[3.0, 10.1]
very high	VH	[5.0, 10.1]
3P%		
语言值	符号	值范围
very low	VL	[0.0, 0.0]

Continuation of Input Variable Scope

low	L	[0.0, 0.2568000000000001]
medium	M	[0.22119999999999998, 0.3724]
high	H	[0.346, 1.0]
very high	VH	[0.4, 1.0]
	2P	
语言值	符号	值范围
very low	VL	[0.0, 0.4]
low	L	[0.0, 1.0]
medium	M	[0.8, 2.2]
high	H	[1.8, 11.2]
very high	VH	[3.6600000000000024, 11.2]
	2PA	
语言值	符号	值范围
very low	VL	[0.0, 0.8]
low	L	[0.0, 2.0]
medium	M	[1.7, 4.44]
high	H	[3.2, 20.3]
very high	VH	[6.8, 20.3]
	2P%	
语言值	符号	值范围
very low	VL	[0.0, 0.39120000000000005]
low	L	[0.0, 0.4822]
medium	M	[0.4646, 0.5608000000000001]
high	H	[0.5376000000000001, 1.0]
very high	VH	[0.6438, 1.0]
	eFG%	
语言值	符号	值范围
very low	VL	[0.0, 0.4122]
low	L	[0.0, 0.4988]
medium	M	[0.47979999999999995, 0.5594]
high	H	[0.5468000000000001, 1.5]
very high	VH	[0.6138, 1.5]
	FT	
语言值	符号	值范围
very low	VL	[0.0, 0.1]

Continuation of Input Variable Scope

low	L	[0.0, 0.5]
medium	M	[0.4, 1.4]
high	H	[1.2, 8.5]
very high	VH	[2.360000000000002, 8.5]
FTA		
语言值	符号	值范围
very low	VL	[0.0, 0.2]
low	L	[0.0, 0.7]
medium	M	[0.6, 1.8]
high	H	[1.5, 11.2]
very high	VH	[2.9800000000000013, 11.2]
FT%		
语言值	符号	值范围
very low	VL	[0.0, 0.1956000000000002]
low	L	[0.0, 0.682]
medium	M	[0.6593999999999999, 0.8161999999999999]
high	H	[0.8, 1.0]
very high	VH	[0.889, 1.0]
ORB		
语言值	符号	值范围
very low	VL	[0.0, 0.1]
low	L	[0.0, 0.4]
medium	M	[0.3, 0.8]
high	H	[0.7, 5.5]
very high	VH	[1.3, 5.5]
DRB		
语言值	符号	值范围
very low	VL	[0.0, 0.8]
low	L	[0.0, 1.7]
medium	M	[1.4, 3.2]
high	H	[2.8, 12.0]
very high	VH	[3.960000000000002, 12.0]
TRB		
语言值	符号	值范围
very low	VL	[0.0, 1.0]

Continuation of Input Variable Scope

low	L	[0.0, 2.1400000000000006]
medium	M	[1.8, 4.0]
high	H	[3.4, 14.3]
very high	VH	[5.4, 14.3]
AST		
语言值	符号	值范围
very low	VL	[0.0, 0.3]
low	L	[0.0, 0.7]
medium	M	[0.6, 1.8]
high	H	[1.5, 9.8]
very high	VH	[3.0, 9.8]
STL		
语言值	符号	值范围
very low	VL	[0.0, 0.1]
low	L	[0.0, 0.3]
medium	M	[0.3, 0.8]
high	H	[0.7, 2.1]
very high	VH	[1.0, 2.1]
BLK		
语言值	符号	值范围
very low	VL	[0.0, 0.0]
low	L	[0.0, 0.2]
medium	M	[0.1, 0.3400000000000006]
high	H	[0.3, 2.5]
very high	VH	[0.7, 2.5]
TOV		
语言值	符号	值范围
very low	VL	[0.0, 0.2]
low	L	[0.0, 0.6]
medium	M	[0.5, 1.1]
high	H	[1.0, 6.2]
very high	VH	[1.680000000000001, 6.2]
PF		
语言值	符号	值范围
very low	VL	[0.0, 0.5200000000000002]

Continuation of Input Variable Scope

low	L	[0.0, 1.4]
medium	M	[1.2, 2.4]
high	H	[2.2600000000000025, 4.7]
very high	VH	[2.9, 4.7]
PTS		
语言值	符号	值范围
very low	VL	[0.0, 1.8200000000000003]
low	L	[0.0, 4.24]
medium	M	[3.8, 9.840000000000003]
high	H	[8.760000000000002, 31.7]
very high	VH	[14.5, 31.7]

表 3: input variable scope

6.3 部分代码

代码结构

```
code
├── source ... 数据
│   ├── 2021-2022 NBA Player Stats-Playoffs.csv ... 源数据
│   ├── players_input ... 各球员处理后的单独信息
│   ├── source_process.py ... 源数据处理
│   └── source_analysis.ipynb ... 源数据分析
├── NBAExpert ... 模糊专家系统
│   ├── bin
│   ├── lib
│   ├── input ... link to code/source/players_input
│   ├── NBA.fcl ... 定义了模糊规则
│   ├── result ... 经过模糊系统处理后的各球员信息
│   └── src
│       └── NBA.java ... 运行模糊系统
└── website
    ├── web.py ... 网页运行
    ├── src ... link to code/NBAExpert/result
    ├── static
    └── templates
```

```
/*
    模糊规则部分的代码
    完整代码参见code/NBAExpert/NBA.fcl
*/

RULEBLOCK defensive
  AND : MIN;  // Use 'min' for 'and'
  ACT : MIN;  // Use 'min' activation method
  ACCU : MAX; // Use 'max' accumulation method


    RULE 1 : IF DRB IS VL THEN defensive IS VB;
    RULE 2 : IF DRB IS L THEN defensive IS B;
    RULE 3 : IF DRB IS M THEN defensive IS M;
  RULE 4 : IF DRB IS H THEN defensive IS G;
    RULE 5 : IF DRB IS VH THEN defensive IS VG;

    RULE 6 : IF DRB IS VL THEN defensive IS VB;
    RULE 7 : IF DRB IS L THEN defensive IS VB;
    RULE 8 : IF DRB IS M THEN defensive IS M;
    RULE 9 : IF DRB IS H THEN defensive IS G;
    RULE 10 : IF TRB IS VH THEN defensive IS G;

    RULE 11 : IF STL IS H THEN defensive IS VG;
    RULE 12 : IF STL IS VH THEN defensive IS VG;

    RULE 13 : IF BLK IS VL THEN defensive IS VB;
    RULE 14 : IF BLK IS L THEN defensive IS B;
    RULE 15 : IF BLK IS M THEN defensive IS M;
    RULE 16 : IF BLK IS H THEN defensive IS VG;
    RULE 17 : IF BLK IS VH THEN defensive IS VG;

END_RULEBLOCK

RULEBLOCK shooting
  AND : MIN;  // Use 'min' for 'and'
  ACT : MIN;  // Use 'min' activation method
  ACCU : MAX; // Use 'max' accumulation method

  RULE 1 : IF FGp IS VL THEN shooting IS VB;
    RULE 2 : IF FGp IS L THEN shooting IS B;
    RULE 3 : IF FGp IS M THEN shooting IS M;
    RULE 4 : IF FGp IS H THEN shooting IS G;
    RULE 5 : IF FGp IS VH THEN shooting IS VG;

```

```
        RULE 6 : IF twoPp IS VL AND threePp IS VL THEN shooting IS VB;
        RULE 7 : IF twoPp IS L AND threePp IS L THEN shooting IS B;
        RULE 8 : IF twoPp IS M AND threePp IS L THEN shooting IS M;
        RULE 9 : IF twoPp IS L AND threePp IS M THEN shooting IS M;
        RULE 10 : IF twoPp IS M AND threePp IS M THEN shooting IS G;
        RULE 11 : IF twoPp IS H OR threePp IS H THEN shooting IS VG;
        RULE 12 : IF twoPp IS VH OR threePp IS VH THEN shooting IS VG;

        RULE 13 : IF FTp IS VL AND threePp IS VL THEN shooting IS VB;
        RULE 14 : IF FTp IS L AND threePp IS L THEN shooting IS B;

    END_RULEBLOCK

    RULEBLOCK rebounding
      AND : MIN;  // Use 'min' for 'and'
      ACT : MIN;  // Use 'min' activation method
      ACCU : MAX; // Use 'max' accumulation method

      RULE 1 : IF TRB IS VL THEN rebounding IS VB;
        RULE 2 : IF TRB IS L THEN rebounding IS B;
        RULE 3 : IF TRB IS M THEN rebounding IS M;
        RULE 4 : IF TRB IS H THEN rebounding IS G;
        RULE 5 : IF TRB IS VH THEN rebounding IS VG;

        RULE 6 : IF ORB IS VL AND DRB IS VL THEN rebounding IS VB;
        RULE 7 : IF ORB IS L AND DRB IS L THEN rebounding IS B;
        RULE 8 : IF ORB IS M AND DRB IS L THEN rebounding IS M;
        RULE 9 : IF ORB IS L AND DRB IS M THEN rebounding IS M;
        RULE 10 : IF ORB IS M AND DRB IS M THEN rebounding IS M;
        RULE 11 : IF ORB IS H OR DRB IS H THEN rebounding IS G;
        RULE 12 : IF ORB IS VH OR DRB IS VH THEN rebounding IS VG;

    END_RULEBLOCK

    RULEBLOCK leading
      AND : MIN;  // Use 'min' for 'and'
      ACT : MIN;  // Use 'min' activation method
      ACCU : MAX; // Use 'max' accumulation method

      RULE 1 : IF AST IS VL THEN leading IS VB;
        RULE 2 : IF AST IS L THEN leading IS B;
        RULE 3 : IF AST IS M THEN leading IS M;
        RULE 4 : IF AST IS H THEN leading IS G;
        RULE 5 : IF AST IS VH THEN leading IS VG;

```

```
        RULE 6 : IF TOV IS VL THEN leading IS VG;
        RULE 7 : IF TOV IS L THEN leading IS G;
        RULE 8 : IF TOV IS M THEN leading IS M;
        RULE 9 : IF TOV IS H THEN leading IS B;
        RULE 10 : IF TOV IS VH THEN leading IS VB;

        RULE 11 : IF STL IS VH OR STL IS H THEN leading IS G;
        RULE 12 : IF ORB IS VH OR ORB IS H THEN leading IS G;
        RULE 13 : IF DRB IS VH THEN leading IS G;

    END_RULEBLOCK

    RULEBLOCK breaking
      AND : MIN;  // Use 'min' for 'and'
      ACT : MIN;  // Use 'min' activation method
      ACCU : MAX; // Use 'max' accumulation method

        RULE 2 : IF FTA IS VL THEN breaking IS VB;
        RULE 3 : IF FTA IS L THEN breaking IS B;
        RULE 4 : IF FTA IS M THEN breaking IS M;
        RULE 5 : IF FTA IS H THEN breaking IS G;
        RULE 1 : IF FTA IS VH THEN breaking IS VG;

        RULE 2 : IF STL IS VH or STL IS H THEN breaking IS G;

        RULE 3 : IF AST IS VH THEN breaking IS VG;
        RULE 4 : IF AST IS H THEN breaking IS G;

        RULE 6 : IF TOV IS VL THEN leading IS G;
        RULE 7 : IF TOV IS L THEN leading IS G;
        RULE 8 : IF TOV IS M THEN leading IS M;
        RULE 9 : IF TOV IS H THEN leading IS B;
        RULE 10 : IF TOV IS VH THEN leading IS VB;

    END_RULEBLOCK
```

References

[1] Pablo Cingolani and Jesus Alcala-Fdez. jfuzzylogic: a robust and flexible fuzzy-logic inference system language implementation. In *2012 IEEE International Conference on Fuzzy Systems*, pages 1–8. IEEE, 2012.

[2] Pablo Cingolani and Jesús Alcalá-Fdez. jfuzzylogic: a java library to design fuzzy logic controllers according to the standard for fuzzy control programming. *International Journal of Computational Intelligence Systems*, 6(sup1):61–75, 2013.

[3] Michael Tiegelkamp and Karl-Heinz John. *IEC 61131-3: Programming industrial automation systems*. Springer, 2010.

[4] Vivo Vinco. 2021-2022 nba player stats - playoffs.csv, version 20, June 4, 2022. Retrieved June 5, 2022 from https://www.kaggle.com/datasets/vivovinco/nba-player-stats.

[5] Rookie road basketball. https://www.rookieroad.com/basketball/. Accessed: June 7, 2022.

[6] Basketball reference. https://www.basketball-reference.com/. Accessed: June 7, 2022.

1 简介

Hopfield Network(HN) 是 John Hopfield 在 1982 年介绍的一种 recurrent artificial neural network[1], 其主要功能是通过定义 neurons 和之间的关系矩阵，从而保存给定的一系列原模式 (predefined patterns)，然后通过定义的 Hopfield Network，还原一个带有噪点的数据，即将模糊数据还原到原数据。

2 优点

作为一个上世纪 80 年代发明的神经网络，到现在已经非常过时，所以我们的优点是着眼于当时的人工智能环境下的分析。

创新性 Hopfield Network 的首要优点我认为是创新性。与之前的单层、多层，使用感知器、反向传播的神经网络不同，Hopfield Network 是一种 recurrent neural network，是之后 Boltzmann Machines、Deep Belief Networks 等更现代神经网络的基础，提供了一种完全不同的神经网络构建角度。它的结构是把所有的神经元全连接，每次迭代的输入是由上次迭代中其他神经元的结果，模拟人脑的联想记忆，是第一批 recurrent neural network。

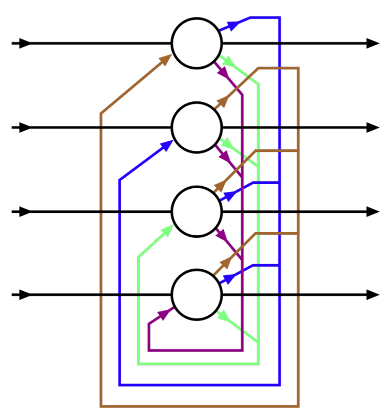


Figure 1: Hopfield Network Structure

稳定性 正如书中所述，上世纪六七十年代许多研究者都在研究循环网络的稳定性，这个问题在 Hopfield Network 中得到解决。

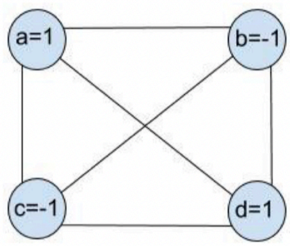
Hopfield Network 的稳定性，或者说其收敛性隐藏在其势能方程中。首先我们有如下定义 Hopfield Network 的方式 [2]。V 表示 predefined pattern，W 表示关系矩阵。那么有势能方程 (1)

$$E = -\frac{1}{2}\sum_i\sum_j w_{ij}V_iV_j = -\frac{1}{2}V^TWV \tag{1}$$

由于神经元每次变化是从-1 到 +1 或者相反，但计算两次迭代势能差 (2) 都是为负，即在迭代中势能是单调下降的。

$$
\begin{aligned}
\Delta E_i &= -\Delta V_i \sum_{j \neq i} w_{ij} V_j \\
&= -\Delta V_i x_i \\
&= -\Delta V_i \cdot W[i,;] \cdot V_i
\end{aligned} \tag{2}
$$

然后可以证明势能会下降到一个固定值，即多次迭代后 ΔE_i 为 0。这等效于 ΔV_i 为 0，当迭代到原始数据后，神经元的值就不会改变，从而达到稳定。



$$
V = \begin{bmatrix} +1 \\ -1 \\ -1 \\ +1 \end{bmatrix}, W = \begin{bmatrix} w_{aa} & w_{ab} & w_{ac} & w_{ad} \\ w_{ba} & w_{bb} & w_{bc} & w_{bd} \\ w_{ca} & w_{cb} & w_{cc} & w_{cd} \\ w_{da} & w_{db} & w_{dc} & w_{dd} \end{bmatrix}
$$

Figure 2: Definition of a Hopfield Network[2]

3 缺点

容量小 如书中所述，最多 (M_{max}) 能存储在 n 神经元 Hopfield Network 中的 predefined patterns 是:

$$M_{max} = 0.15n$$

想要几乎完全完美检索所有 patterns，记忆数减为:

$$M_{max} = \frac{n}{2\ln n}$$

而想要完全检索所有 patterns，记忆数还会减半:

$$M_{max} = \frac{n}{4\ln n}$$

在之后的研究中，将书中描述的 Hopfield Network 称为离散型的 HN，其神经元只能作两极变换，所以存储信息较少，而后研究出了连续型的 HN，从而储存了指数级的记忆信息 [4]，改进了这一点。

Local Minima Problem 即书中描述的得出错误结果的例子，Hopfield Network 的势能下降可能会掉进一个区域最小值 [3]，得出错误结论。

Discrepancy Limitation 在 Hopfield Network 中，要求输入的数据和期望得到的数据差距不过 25%[3]，这是由 Hopfield Network 迭代方式决定的，将 HN 的状态想象成多维立方体，每次迭代就是沿一条边移动，那么当 predefined patterned 都在对角位置上是，识别区分度最好，从而有数据差距的限制。

记忆模式 Hopfield Network 是发展出来一种模仿人类神经活动的神经网络，参照了人类可以从一个不完整的记忆，还原完整记忆的形式。但由于 Hopfield Network 单层网络的缘故，不能模仿人类从一种记忆联想另一种记忆的能力，针对这单点发展出了后续的双向联想记忆。

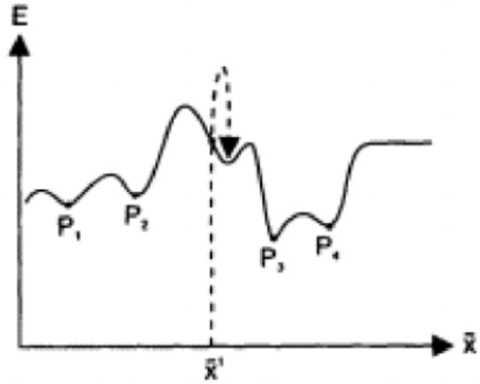


Figure 3: Local Minima Problem[3]

References

[1] John J Hopfield. "Neural networks and physical systems with emergent collective computational abilities." In: *Proceedings of the national academy of sciences* 79.8 (1982), pp. 2554–2558.

[2] *Hopfield Networks are useless.* `https://towardsdatascience.com/hopfield-networks-are-useless-heres-why-you-should-learn-them-f0930ebeadcd`. Accessed: June 25, 2022.

[3] Nikola K Kasabov. *Foundations of neural networks, fuzzy systems, and knowledge engineering.* Marcel Alencar, 1996.

[4] Hubert Ramsauer et al. "Hopfield networks is all you need". In: *arXiv preprint arXiv:2008.02217* (2020).