

# 基于加权网络和局部适应度的蛋白质复合物识别算法\*

刘翠翠<sup>1</sup>, 孙伟<sup>2</sup>

(1. 长沙医学院 信息工程学院, 长沙 410219; 2. 信息工程大学 网络空间安全学院, 郑州 450000)

**摘要:** 通常在蛋白质网络中挖掘稠密子图或模块来识别其中的蛋白质复合物, 限制了其应用范围和识别的准确性。针对该问题, 提出了一种基于加权网络和局部适应度的蛋白质复合物识别算法。该算法综合稠密子图的密度指标和模块性定义了新的局部适应度函数, 并基于边聚集系数构建加权的蛋白质网络, 根据权值选择边, 在加权蛋白质网络中将种子边不断聚类扩展, 挖掘综合适应度最大的子图, 从而识别出蛋白质复合物。在多个真实蛋白质网络中的实验表明, 该算法能够有效提升蛋白质复合物识别的准确性。

**关键词:** 加权网络; 适应度; 蛋白质复合物识别; 模块性

**中图分类号:** TP391

**文献标志码:** A

**文章编号:** 1001-3695(2018)08-2308-03

**doi:** 10.3969/j.issn.1001-3695.2018.08.017

## Algorithm for identifying protein complexes based on weighted network and local fitness

Liu Cuicui<sup>1</sup>, Sun Wei<sup>2</sup>

(1. College of Information Engineering, Changsha Medical University, Changsha 410219, China; 2. Institute for Network Security, Information Engineering University, Zhengzhou 450000, China)

**Abstract:** Usually, the ways about mining dense subgraph or module would limit their scope of application and recognition accuracy on protein complexes identification. To solve this problem, this paper proposed a novel protein complex recognition algorithm based on weighted network and local fitness. By integrating density of subgraph and modularity, it defined a new local fitness function, and used edge clustering coefficient to construct the weighted protein network, selected some seed edges according to their weights, then extended the clustering around the seed edge until gaining a biggest protein subgraph with maximum comprehensive fitness. Experiments in yeast protein networks show that, this algorithm can effectively improve the accuracy on protein complexes identification.

**Key words:** weighted networks; fitness; protein complexes identification; modularity

## 0 引言

蛋白质复合物是组成蛋白质的基础结构, 大部分的生物过程都需要蛋白质复合物参与并实现。因此, 有效地识别蛋白质复合物对理解高分子生物作用等都具有重要意义<sup>[1,2]</sup>。当前, 高通量技术已经累积了大量的蛋白质数据; 同时, 基于复杂网络的模块(社区)发现或聚类算法也开始应用于蛋白质复合物识别, 这些都为网络级蛋白质复合物的挖掘奠定了基础。随着生物信息技术的发展, 越来越多的数据挖掘、聚类算法被应用于识别和分析蛋白质分子结构<sup>[3]</sup>, 并从揭示复杂系统内部结构功能的角度, 引导和提升蛋白质复合物识别的效率。

2003年, Spirin等人<sup>[4]</sup>发现蛋白质复合物往往由蛋白质网络中的子团或子图构成。据此, CPM/MCODEE等算法提出, 通过挖掘子团或子图以发现蛋白质复合物。近年来研究发现, 不在稠密子图范围内的蛋白质, 也可能属于蛋白质复合物。进一步地, 人们发现很多复合物并不属于稠密子图<sup>[3]</sup>。为了解决这个问题, 即使是非稠密子图的蛋白质复合物, 其内部连接也比外部要更加紧密<sup>[1]</sup>。由此定义了模块性(modularity)指标, 通过挖掘蛋白质模块以识别复合物。这类算法虽然能识别各种不同密度的复合物, 但对于模块度不高的复合物识别准确率不高。

比较而言, 基于模块的识别算法容易忽略内聚性较低模块度不高的蛋白质复合物, 而基于稠密子图的算法容易忽略低密度的复合物。针对两类算法的优缺点, 本文提出了一种基于加权网络和局部适应度的识别算法。一方面, 该算法综合稠密子图的密度指标和模块性定义了新的局部适应度函数; 另一方面, 该算法根据边聚集系数值构建加权的蛋白质网络, 并根据

权值选择边, 围绕种子边不断聚类, 选取最大综合适应度的子图作为识别的复合物。在多个真实的蛋白质网络中的实验表明, 该算法能够有效提升蛋白质复合物识别的准确性。

## 1 问题描述

现有的研究表明<sup>[5,6]</sup>, 蛋白质复合物往往表现为网络中的稠密子图或模块。然而在实际应用中, 基于模块性的识别算法难以发现低模块性的复合物, 基于稠密子图的识别算法无法挖掘低密度的复合物。因此, 只有同时考虑子图密度和模块度, 才能保证蛋白质复合物识别的准确性。基于加权网络和局部适应度的蛋白质复合物识别算法的基本思想为: 首先基于蛋白质网络构建加权网络, 从种子边出发, 根据邻居节点的局部适应度逐步扩展, 从而构建该种子边所在的蛋白质复合物, 最终获得蛋白质网络中所有复合物的结构分布。

本文综合子图密度与模块性两种特征提出了局部适应度指标, 并将复合物识别建模为寻找局部适应度最大的蛋白质子图。为度量蛋白质网络节点的局部适应度, 这里首先给出相关定义。

### 1.1 相关定义

**定义1** 子图密度。令  $G = \{V, E, \omega\}$  表示一个蛋白质网络, 其中:  $V$  是蛋白质网络节点集;  $E$  是蛋白质网络边集;  $\omega$  表示网络连接权重。子图  $P$  在网络  $G$  中包含  $P_e$  条边、 $P_v$  个节点, 则定义子图密度为

$$D_P = \frac{2 \times P_e}{P_v(P_v - 1)} \quad (1)$$

**定义2** 子图模块度。令  $G = \{V, E, \omega\}$  表示一个蛋白质

收稿日期: 2017-04-12; 修回日期: 2017-05-16 基金项目: 国家自然科学基金资助项目(F010103)

作者简介: 刘翠翠(1983-), 女, 湖南岳阳人, 讲师, 硕士, 主要研究方向为生物信息学(liucucui1983@21cn.com); 孙伟(1971-), 男, 河南郑州人, 讲师, 硕士, 主要研究方向为网络信息安全。

网络,其中: $V$ 是蛋白质网络节点集; $E$ 是蛋白质网络边集; $\omega$ 表示网络连接权重。子图 $P$ 在网络 $G$ 中包含 $P_e$ 条边、 $P_v$ 个节点,则定义子图模块度为

$$M_P = \frac{\sum_{v \in P} d_{in}(P, v)}{\sum_{v \in P} d_{out}(P, v) + \sum_{v \in P} d_{in}(P, v)} \quad (2)$$

其中: $v$ 是子图 $P$ 中任意节点; $d_{in}(P, v)$ 表示子图 $P$ 内的节点 $v$ 的加权内度,是节点 $v$ 与子图 $P$ 内部其他节点的连接边的权重之和;同样地, $d_{out}(P, v)$ 表示子图 $P$ 内的节点 $v$ 的加权外度,是节点 $v$ 与子图 $P$ 外部其他节点的连接边的权重之和;模块度 $M_P$ 实际上表示子图 $P$ 的加权内度与所有度数之比,能够反映该子图的内聚特性。

基于子图密度和子图模块度的定义,可进一步定义子图适应度。

**定义3** 子图适应度。给定一个蛋白质网络 $G = \{V, E, \omega\}$ 和子图 $P(P \subseteq G)$ ,定义子图适应度为

$$f_P = D_P M_P^{(1-\rho)} \quad (3)$$

其中: $\rho \in (0, 1)$ 表示子图密度与模块度两种特征在 $f_P$ 中的重要程度。 $\rho$ 越大,说明子图密度在 $f_P$ 中的影响力越大,而子图模块度的影响力越小;反之,则子图密度的影响力越小,而子图模块度的重要性越高。

**定义4** 节点适应度。给定一个蛋白质网络 $G = \{V, E, \omega\}$ 和子图 $P(P \subseteq G)$ ,对于网络中任意节点 $v$ 关于子图 $P$ 的节点适应度表示为

$$f(P, v) = f_{P+v} - f_{P-v} \quad (4)$$

其中: $P+v$ 与 $P-v$ 分别表示在子图 $P$ 中加入节点 $v$ 和删除节点 $v$ 的聚类选择策略; $f_{P+v}$ 和 $f_{P-v}$ 分别表示 $P+v$ 与 $P-v$ 时的子图适应度值; $f(P, v)$ 表示子图 $P$ 在包含节点 $v$ 和不包含节点 $v$ 时子图适应度的差值。式(4)表示节点 $v$ 对于子图 $P$ 的适应度贡献值。当 $f(P, v)$ 越大,则节点 $v$ 越可能属于子图 $P$ 。

## 1.2 基于加权网络和局部适应度的算法识别模型

假设蛋白质网络中具有最大适应度的子图即为蛋白质复合物。根据本文的识别算法,初始化阶段从种子边向外扩展,综合考虑子图密度与模块度指标融合邻近的节点,直至最后获取有最大适应度的子图。如图1所示,首先从种子边出发,计算节点的局部适应度,根据适应度确定节点的聚类与删除。在聚类过程中,当某邻居节点能增加适应度时,则选择融合该节点;当某节点降低适应度时,则选择删除该节点。通过迭代搜索遍历网络中的蛋白质节点,使得最终获得的子图局部适应度最大。

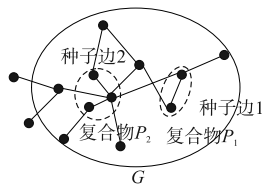


图1 蛋白质复合物识别算法模型

与传统的复合物识别算法相比,本文算法具有两项特点:a)适应度指标综合考虑了子图密度、模块度的双重影响,有助于提升复合物识别的精确性;b)聚类过程中子图可以删除、加入节点,甚至合并子图,能够较为精确地控制复合物的聚类过程和节点归属。

## 2 基于加权网络和局部适应度的蛋白质复合物识别算法

### 2.1 加权网络构造与选择种子边

若给定的蛋白质网络为加权网络,并且其边权值能够反映蛋白质网络中节点的相互作用和联系,则可以根据蛋白质网络的边权值直接选择种子边;若给定的蛋白质网络为非加权网络,或该网络的边权值没有反映蛋白质网络中节点的作用和联

系,比如边权值表示相互作用的可信度,根据相关研究,采用文献[7]提供的方法,通过网络预处理将非加权的蛋白质网络改造为加权网络,从而使得网络中的边权值能够反映蛋白质网络中节点的相互作用和联系,以便于后续的蛋白质复合物识别。

有研究发现<sup>[8,9]</sup>,在蛋白质网络中,边聚集系数越高的相互作用越可能定位蛋白质复合物。边聚集系数既可应用于加权网络,又可应用于非加权网络。因此,本文根据边聚集系数选取种子边集合,并围绕核心边聚类进行蛋白质复合物识别。

一般而言,蛋白质网络中边聚集系数较高的节点,更有可能围绕该节点扩展为蛋白质复合物。因此,蛋白质网络中复合物发现需要在初始化阶段获取核心的种子边,以提升识别算法的准确性和有效性。获取核心种子边的思路如下:选取蛋白质网络中节点度高于均值的所有节点,计算这些边的边聚集系数,获取满足条件的边作为种子边。基于此,本文提出一种简单有效的种子边选择策略。其步骤如下:

- 初始化集 $S_q$ 为空;
- 获取蛋白质网络中所有节点度高于均值的节点集合,遍历这些节点的连接边;
- 当 $E_{clu}(i) \geq \tau$ ,即边 $i$ 的边聚集系数大于或等于相应的门限,则将 $i$ 增加到集合 $S_q$ ;
- 重复步骤b)c),直至所有满足边聚集系数条件的种子边为止。

### 2.2 蛋白质复合物识别算法

针对两类算法的优缺点,本文提出了一种基于加权网络和局部适应度的蛋白质复合物识别算法。一方面,该算法综合稠密子图的密度指标和模块性定义了新的局部适应度函数;另一方面,该算法根据边聚集系数值构建加权的蛋白质网络,并根据权值选择边,在蛋白质网络中围绕种子边不断聚类,并获取最大适应度的子图作为被识别的复合物。该算法的具体步骤如下:

- 初始化种子边集合 $S_q$ 、蛋白质复合物集合 $C$ 为空集。
- 计算边聚集系数,采用种子边选择算法(FindCoreEdges)获取核心种子边集合 $S_q$ 。
- 将集合 $S_q$ 中的种子边按照降序排列,围绕种子边进行子图扩张,获取各个种子边所在的子图。为控制计算复杂度,种子边的边聚集系数须大于门限阈值 $\xi$ 。
- 根据式(4)计算节点 $v_i$ 对于子图 $Clu$ 的节点适应度。若 $f(Cl_u, v_i) \geq 0$ ,则将该节点加入子图 $Cl_u$ 。
- 根据式(4)计算邻居子图 $p$ 与给定种子边所属的节点子图 $Cl_u$ 的适应度。若 $f(Cl_u, p) \geq 0$ ,合并两节点子图,得到更新后的 $Cl_u$ 。
- 重复步骤d)e),以遍历所有的种子边和子图,得到蛋白质网络中所有的蛋白质复合物集合 $C$ 。

根据上述步骤说明,本文算法描述如下:

**算法1** 基于加权网络和局部适应度的蛋白质复合物识别算法

```

输入:  $G(V, E, \omega)$ 。
输出: 蛋白质网络中所有的蛋白质复合物  $C$ 。
Initialize  $C, S_q$ ;
 $S_q \leftarrow \text{FindCoreEdges}()$ ; // 获取蛋白质网络中种子边集合  $S_q$ 
% 获取所有种子边所在的子图 %
for each  $i \in N$  do
    Initialize  $LC(i) \leftarrow i$ ; // 初始化核心子团  $LC(i)$  为核心节点
    for each  $\{j | \text{sim}(j, LC(i)) \geq \xi\}$  do
        calculate  $\text{sim}(j, LC(i))$  by Eq. 2;
        calculate  $N_{LC(i)}^j$  by Eq. 5;
        if  $N_{LC(i)}^j > 0$ 
             $N_{LC(i)}^j \leftarrow LC(i) \cup \{j\}$ ; // 将节点  $j$  加入到  $LC(i)$ 
        end if
    end for
end for
end for

```

```

% 蛋白质网络中复合物识别%
for each  $|LC(i) \setminus LC(i) \cap LC(s) \neq \emptyset|$  do
    if  $\text{sim}(LC(i), LC(s)) \geq \xi$ 
         $LC(s) \leftarrow LC(i) \cup LC(s)$ ; // 将  $LC(i)$  加入到  $LC(s)$ 
    end if
end for
return  $Clu$ .

```

### 3 实验与结果分析

为了分析基于加权网络和局部适应度的蛋白质复合物识别算法的性能,实验将其同时与四种经典的蛋白质复合物识别算法进行比较,即 CPM<sup>[10]</sup>、CMC<sup>[11]</sup>、MCL<sup>[12]</sup> 和 NFC<sup>[13]</sup> 算法。其中,CPM、CMC 是典型的基于稠密子图的蛋白质复合物识别算法,MCL 则是一种基于随机流应用网络聚类的高效挖掘算法。NFC 则是一种基于模块度的局部适应性挖掘算法。后两种算法都可以视为基于模块的复合物挖掘算法。四种算法的参数均采用其文献提供的最佳参数值,具体配置如表 1 所示。

表 1 四种算法参数配置

算法	参数设定
本文	$\rho = 0.2$
CMC	$\text{adjust} = 1, \text{overlap}_{\text{thres}} = 0.5$ $\text{merge}_{\text{thres}} = 0.15, \text{size} = 0.3$
CPM	$k = 3$
MCL	$\text{inflation} = 2.0$
NFC	$\alpha = 1$

实验中采用的蛋白质网络源于 DIP 数据库中的酵母蛋白质网络和大肠杆菌蛋白质网络<sup>[14]</sup>。在删除自相互作用和重复相互作用后统计发现:酵母蛋白质网络包含 4 746 个蛋白质和 15 166 条作用边;大肠杆菌蛋白质网络包含 2 727 个蛋白质和 11 803 条作用边。由于这两种网络均为非加权网络,可使用文献[8]推荐的方法进行预处理,使之转换为权重网络。将算法识别结果与已知的蛋白质网络相比较,从而直接反映算法对蛋白质复合物识别的有效性和转缺陷。

一般而言,蛋白质网络中被标志的复合物比例越高,识别得越精确,说明算法识别蛋白质复合物的能力越强。其中,特异性 (specificity,  $Sp$ ) 表示所识别的蛋白质复合物中识别正确部分的比重,而敏感度 (sensitivity,  $Sn$ ) 表示已知蛋白质复合物中被算法标志出来的部分所占比重。

$$Sp = \frac{TP}{TP + FP} \quad (5)$$

$$Sn = \frac{TP}{TP + FN} \quad (6)$$

其中: $FP$  表示蛋白质网络中未被标志为已知的复合物个数; $TP$  表示蛋白质网络中标志为已知复合物的个数; $FN$  表示已知复合物中未被算法识别的个数。 $Sp$  与  $Sn$  各有侧重,为综合这两项指标,Wang 等人提出了综合评价指标 ( $F$ -score,  $F$ )。

$$F = \frac{2 \times Sp \times Sn}{Sp + Sn} \quad (7)$$

$F$  指标融合了已知复合物的识别率及复合物识别的正确率两项性能指标,从而保证这两项指标不受蛋白质网络规模、复合物识别个数的影响。因此,实验中采用了综合评价  $F$  来评估算法识别蛋白质复合物的性能。

为全面比较算法的识别性能,本文算法与同类算法就各种算法的  $F$ -score、特异性 ( $Sp$ )、敏感度 ( $Sn$ ) 和识别复合物的匹配数量分别进行比较,如表 2 所示。

表 2 四种算法的蛋白质复合物识别的性能对比

算法	识别数量	$Sn$	$Sp$	$F$ -score	完全匹配数量
CPM	197	0.185	0.376	0.248	11 (5.58%)
CMC	1 130	0.576	0.219	0.317	21 (1.86%)
MCL	929	0.450	0.187	0.264	15 (1.61%)
NFC	518	0.277	0.209	0.238	5 (0.97%)
本文算法	297	0.452	0.630	0.526	28 (9.43%)

就单一项指标而言,本文算法识别的蛋白质复合物个数为

297 个,有三种算法的识别数量要大大超过它;就蛋白质识别的敏感度而言,本文算法也弱于 CMC 算法。但综合蛋白质复合物识别准确率的五项指标,蛋白质识别的特异性、综合指标  $F$ -score 都是最优的。更重要的是,本文算法识别的完全匹配的蛋白质复合物数量最多,在其识别数量中的比例最高。这说明与其他蛋白质复合物识别算法相比,本文算法在蛋白质复合物完全匹配识别、综合识别准确率方面都具有更大的优势,识别蛋白质复合物的性能更优。

如图 2 所示,对于已知低模块性的蛋白质网络,在不同的匹配阈值下,本文算法蛋白质复合物识别的匹配准确率要远高于其他算法,相比 NFC 算法优势尤为明显。图 2 中,当匹配阈值越低,本文算法在低模块性蛋白质网络中的复合物识别准确率越高;当匹配阈值越高,本文算法在低模块性蛋白质网络中的复合物识别准确率逐步下降,但也具有更高的复合物识别准确率。相比其他四种蛋白质识别算法,本文算法能更有效、更准确地识别低模块性的已知复合物。

图 3 表示五种算法识别已知低密度复合物的性能比较。针对已知低密度蛋白质网络,当匹配阈值不高于 0.7 时,本文算法识别复合物的准确率远高于同类算法,与基于稠密子图的识别算法相比提升尤为明显。图 3 中,当匹配阈值为 0.2 时,本文算法的  $F$ -score 指标为 0.303,与 CPM 算法相比提高了 85%,比 CMC 算法提升了 1.5 倍。综合比较五种蛋白质复合物识别算法,本文算法能更准确地发现已知低密度的蛋白质复合物。

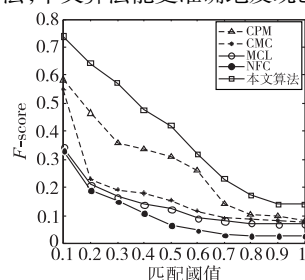


图 2 五种算法识别已知低模块性蛋白质的准确率  $F$ -score 对比

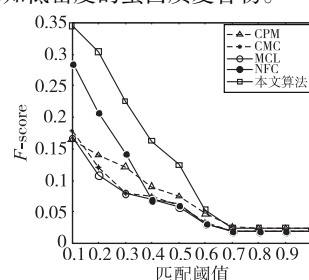


图 3 五种算法识别已知低密度蛋白质的准确率  $F$ -score 对比

综合比较上述三项实验,基于稠密子图与基于模块的复合物识别算法各有其自身难以克服的局限性,无法同时就两项特征(低密度性、模块性)充分、精确地识别蛋白质网络中的复合物。与其他四种蛋白质复合物挖掘算法相比,本文算法在识别具有不同拓扑结构的蛋白质复合物方面,特别是针对低密度、低模块性的蛋白质网络,识别准确性更高,能够大大优化蛋白质复合物的识别结果。

### 4 结束语

蛋白质复合物结构复杂,子图结构和模块性都属于其结构特征。仅仅依靠某一种特征进行挖掘,无疑会影响复合物识别的效率和准确性。本文提出了一种基于加权网络和局部适应度的复合物挖掘算法,融合了模块性与子图密度特征识别,以充分、精确地识别蛋白质网络中具有复杂结构特征的复合物。本文算法综合稠密子图的密度指标和模块性定义了新的局部适应度函数,并基于边聚集系数构建加权的蛋白质网络,根据权值选择边,在加权蛋白质网络中将种子边不断聚类扩展,挖掘综合适应度最大的子图,从而识别出蛋白质复合物。

为验证本文算法的有效性,将其与其他四种经典算法进行比较。通过与已知蛋白质复合物进行匹配分析,针对具有不同结构特点的蛋白质网络(低密度或低模块性),本文算法在识别准确率和有效性方面都有了较大的性能提升。这说明该算法能够有效提升蛋白质复合物识别的性能,改善蛋白质网络中复合物识别的精确度。

#### 参考文献:

- [1] Barabasi A L, Oltvai Z N. Network biology: understanding the cells' functional organization [J]. *Nature Reviews Genetics*, 2004, 5 (2): 101-113.

(下转第 2327 页)

加障碍的复杂度,如图2所示。由于环境变复杂,所以同时将恒定温度参数的仿真实验和动态温度参数的仿真实验的 Episode 设置为 4 000。在恒定温度参数仿真实验中,如图5(a)所示,PSG-HAQL、H-HAQL 算法的成功率远远高于 L-HAQL、Standard-QL;同样地,如表4所示,H-HAQL 和 L-HAQL 虽然均获得步数为 55 的最优策略,但是 L-HAQL 确定最优策略的回报值要低于 H-HAQL。此外,只有 PSG-HAQL 算法在设定的 Episode 下寻找到全局最优策略。且随着环境的复杂化,PSG-HAQL 算法的性能改善更加明显。在动态温度参数仿真实验中,如表5、图6(b)(c)所示,在较为复杂的环境中,只有 PSG-HAQL 算法在设定 Episode 次训练中能够找到最优策略,其他三种算法虽然较之恒定温度参数仿真实验结果有所改进,但是在设定 Episode 次训练中均未能找到最优策略。其他实验结果与 20×20 仿真环境中的大体相同。

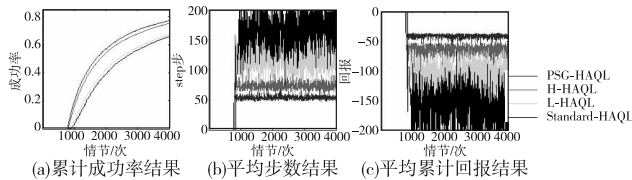


图5 30×30仿真环境—恒定温度参数仿真实验结果

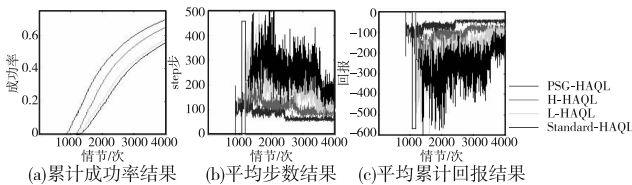


图6 30×30仿真环境—动态温度参数仿真实验结果

通过以上四组仿真对比实验,首先验证了随着每次训练结果优劣情况更新的启发函数能够改善强化学习性能表面。虽然启发函数的影响量级较大时能够使 agent 更早地找到目标位置,但是较大的启发函数影响量级也会使算法陷入局部收敛。采用依据实验结果优劣更新的信息强度,然后由信息强度更新启发函数的影响量级来启发强化学习的 PSG-HAQL 算法能够更早地收敛到最优策略,有效提高强化学习算法的收敛速度。

## 5 结束语

启发式 Q 学习算法可以在线更新启发函数,加快强化学习速度。本文在启发式 Q 学习的基础上,提出了可在线更新的信息强度引导的启发式 Q 学习(PSG-HAQL)算法。在本文中,首先给出了信息强度更新方法和信息强度启发方法的定义,同时应用 Boltzmann 机制进行启发式强化学习,进而给出了 PSG-HAQL 算法;然后证明了 PSG-HAQL 算法的收敛性;最后将 PSG-HAQL 算法应用于路径规划仿真实验以验证其性能。

PSG-HAQL 算法将蜂群信息传递的思想结合到启发式 Q 学习方法:agent 在训练过程中不断获得不同策略的适应度以在线更新该策略信息强度,将信息强度作为 Q 学习启发函数,使 agent 有更高概率去选择信息强度高的策略。所以,信息强度引导的启发式 Q 学习(PSG-HAQL)算法能够更高效地寻找到最优策略,从而进一步缩减训练时间。本文的研究工作主要集中于将蜂群信息思想引入到启发式 Q 学习模型中,并通过仿真实验验证其性能改善;下一步的工作主要是考虑将信息强度引导的启发式强化学习与价值网络相结合,从而在一定程度上提高深度 Q 网络(DQN)的收敛速度。

## 参考文献:

- [1] 王雪松,朱美强,程玉虎. 强化学习原理及其应用[M]. 北京: 科学出版社, 2014: 1-2.
- [2] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. *Nature*, 2015, 518(7540): 529.
- [3] Levine S, Wagener N, Abbeel P. Learning contact-rich manipulation skills with guided policy search[C]//Proc of IEEE International Conference on Robotics and Automation. Piscataway, NJ: IEEE Press, 2015: 156-163.
- [4] 陈学松,杨宜民. 强化学习研究综述[J]. 计算机应用研究, 2010, 27(8): 2834-2838,2844.
- [5] 于俊. 模型无关的贝叶斯强化学习方法研究[D]. 苏州: 苏州大学, 2014.
- [6] 张汝波. 提高强化学习速度的方法研究[J]. 计算机工程与应用, 2001, 37(22): 38-40.
- [7] Torrey L, Shavlik J, Walker T, et al. Skill acquisition via transfer learning and advice taking[M]. Berlin: Springer, 2006: 425-436.
- [8] Bianchi R A, Ribeiro C H, Costa A H. Accelerating autonomous learning by using heuristic selection of actions[J]. *Journal of Heuristics*, 2008, 14(2): 135-168.
- [9] 方敏,李浩. 基于状态回溯代价分析的启发式 Q 学习[J]. 模式识别与人工智能, 2013, 26(9): 838-844.
- [10] Sutton R S, Barto A G. Reinforcement learning: an introduction[M]. 2nd ed. Cambridge, MA: MIT Press, 2016: 138-139.
- [11] 冀俊忠,魏红凯,刘椿年,等. 基于引导素更新和扩散机制的人工蜂群算法[J]. 计算机研究与发展, 2013, 50(9): 2005-2014.
- [12] 胡晓辉. 一种基于动态参数调整的强化学习动作选择机制[J]. 计算机工程与应用, 2008, 44(28): 29-31.
- [13] Bianchi R A C, Ribeiro C H C, Costa A H R. Heuristically accelerated Q-learning: a new approach to speed up reinforcement learning[C]//Advances in Artificial Intelligence. Berlin: Springer, 2004: 245-254.
- [14] 魏英姿,赵明扬. 强化学习算法中启发式回报函数的设计及其收敛性分析[J]. 计算机科学, 2005, 32(3): 190-193.
- [15] 刘全,高阳,陈道蓄,等. 一种基于启发式轮廓表的逻辑强化学习方法[J]. 计算机研究与发展, 2008, 45(11): 1824-1830.

(上接第 2310 页)

- [2] 冀俊忠,刘志军,刘红欣,等. 蛋白质相互作用网络功能模块检测的研究综述[J]. 自动化学报, 2014, 40(4): 577-593.
- [3] Luo Feng, Yang Yunfeng, Chen C F, et al. Modular organization of protein interaction networks[J]. *Bioinformatics*, 2007, 23(2): 207-214.
- [4] Spirin V, Mimi L A. Protein complexes and functional modules in molecular networks[J]. *Proc of the National Academy of Sciences*, 2003, 100(21): 12123-12128.
- [5] 王杰,梁吉业,郑文萍. 一种面向蛋白质复合体检测的图聚类方法[J]. 计算机研究与发展, 2015, 52(8): 1784-1793.
- [6] 胡赛,熊慧军,李学勇,等. 多关系蛋白质网络构建及其应用研究[J]. 自动化学报, 2015, 41(12): 2155-2163.
- [7] 刘阳,季新生,刘彩霞. 网络社区发现优化: 基于随机游走的边权预处理方法[J]. 电子与信息学报, 2013, 35(10): 2335-2340.
- [8] 汪涛,刘阳,席耀一. 一种基于核心节点跳转的局部社区发现算法[J]. 上海交通大学学报, 2015, 49(12): 1809-1816.

- [9] Wang Jianxin, Ren Jun, Li Min, et al. Identification of hierarchical and overlapping functional modules in PPI networks[J]. *IEEE Trans on Nanobioscience*, 2012, 11(4): 386-393.
- [10] Palla G, Derenyi I, Farkas I, et al. Uncovering the overlapping community structure of complex networks in nature and society[J]. *Nature*, 2005, 435(7043): 814-818.
- [11] Liu Guimei, Wong L, Chua H N. Complex discovery from weighted PPI networks[J]. *Bioinformatics*, 2009, 25(15): 1891-1897.
- [12] Iasblom J, Wodak S J. Markov clustering versus affinity propagation for the partitioning of protein interaction graphs[J]. *BMC Bioinformatics*, 2009, 10(1): 99.
- [13] Lancichinetti A, Fortunato S, Kertesz J. Detecting the overlapping and hierarchical community structure in complex networks[J]. *New Journal of Physics*, 2009, 11(3): 033015.
- [14] Xenarios I, Salwinski L, Duan X J, et al. DIP, the database of interacting proteins: a research tool for studying cellular networks of protein interactions[J]. *Nucleic Acids Research*, 2002, 30(1): 303-305.