

文章编号: 1674-8085(2019)04-0040-06

基于热点度轨迹显影机制的网络社区聚类算法

*陶 硕¹, 刘 盈²

(1. 马鞍山职业技术学院电子信息系, 安徽, 马鞍山 243031; 2. 井冈山大学电子与信息工程学院, 江西, 吉安 343009)

摘 要: 为解决网络社区聚类算法在实际应用中存在热点捕捉困难和社区聚类生存时间较低的问题, 提出了一种基于热点度轨迹显影机制的网络社区聚类算法。首先, 考虑网络社区聚类存在的多径一体特性, 采用抽样方式与角度估计方法来实现热点信号的精确捕捉, 以提高聚类效率; 随后, 对热点信号矢量空间进行按列重排, 并综合考虑传输矩阵具有的按列正交及全秩特性, 构建热点度轨迹显影方法, 以提高聚类中热点显影速度和增加聚类生存时间。仿真实验表明: 与聚类流动性映射算法 (Clustering Liquidity Mapping Algorithms, CLM 算法)、超欧里几何热度聚类算法 (Hyper-Eulerian Geometric Thermal Clustering Algorithms, H-EGTC 算法) 相比, 所提算法具有更低的聚合时间和搜寻失误差率, 以及更高的热点显示时间。

关键词: 网络社区; 热点捕捉; 聚类聚合; 热点显影; 角度估计; 传输矩阵

中图分类号: TP393

文献标识码: A

DOI:10.3969/j.issn.1674-8085.2019.04.008

THE CLUSTERING ALGORITHM OF NETWORK COMMUNITY BASED ON HOTSPOT TRAJECTORY DEVELOPMENT MECHANISM

*TAO Shuo¹, LIU Ying²

(1. Department of electronic information, Maanshan Technical College, Maanshan, Anhui 243031, China; 2. School of Electronic and Information Engineering, Jinggangshan University, Ji'an, Jiangxi 343009, China)

Abstract: In order to solve some shortcomings of the clustering algorithm in practice, such as difficulty in quickly catching hot spots, short clustering lifetime and high energy consumption, a clustering algorithm based on hotspot trajectory development mechanism is proposed. Firstly, considering the characteristics of multi-path integration which widely exists in network community clustering, and considering that the current network community clustering algorithms are mostly used in wireless sensor networks and the Internet of Things and other application scenarios, the sampling method and angle estimation method are used to achieve accurate capture of hot spot signals and enhance the efficiency of clustering aggregation process. Then, the hot spot signal vector space is rearranged according to the column for comprehensive consideration. Considering the orthogonal and full rank characteristics of the transfer matrix, a simple sorting mode of hotspot trajectory development method is constructed to improve the speed of hotspot development in clustering and increase the survival time of clustering. The simulation results show that compared with CLM and H-EGTC algorithms, the proposed algorithm has the advantages of less aggregation time, higher hotspot display time, lower search error rate and superior energy consumption. It has strong practical deployment value.

Key words: Network community; hotspot capture; clustering aggregation; hotspot development; angle estimation; transmission matrix

收稿日期: 2019-02-08; 修改日期: 2019-04-30

作者简介: *陶 硕(1973-), 女, 安徽枞阳人, 讲师, 硕士, 主要从事数据挖掘、社区网络分析、计算机应用技术等方面的研究(taoshuo1973@sina.com); 刘 盈(1982-), 女, 江西吉安人, 实验师, 硕士, 主要从事计算机应用技术方面的研究(E-Mail:50411709@qq.com).

0 引言

近年来,网络社区理论已经从传统的社交领域逐步渗透入移动无线传感网(Mobile Wireless Sensor Network, M-WSN)、物联网(Internet of Things, IOT),网络社区聚类算法也开始进行实际应用部署,以解决网络定位、数据传输等网络技术难题^[1-2]。

网络社区聚类算法作为一种能够有效捕捉网络热点并实现对网络节点形态及模式追踪的常用方法,已成为国内外学者的研究热点。如 Kai 等^[3]提出了一种基于时空二向度特征捕捉机制的网络社区聚类算法,该方法通过正交方式在时间、空间两个维度进行聚类特征捕捉,并采用周期时移机制实现热点聚类的动态更新,提高聚类生成及更新的效率,在移动无线传感网领域具有较好的应用场景。但是,该算法实现过程复杂,需要采用复杂的信道-信源信号过滤预成型机制来实现特征构拟,在实际应用中,难以进行大规模的部署应用。Zhi 等^[4]提出了一种基于连通度热点更新机制的网络社区聚类算法,该方案主要通过区域广播算法实现对拓扑联通状况的实时感知,且使用君士坦丁寻址机制来精确获取基于流量交互度的连通评估阈值,有效捕捉节点与周围节点的连通度,实现过程具有便捷化特点。然而,该算法需要采用被动旋跳方式进行区域广播,易造成严重的网络拥塞现象。Chun 等^[5]提出了一种基于区域节点竞争机制的网络社区聚类算法,通过预设区域聚类节点方式进行区域更新,且在文献[4]的基础上针对区域聚类节点的连通度进行二次聚类,具有很强的传输稳定性能。但是,由于聚类形成过程需要采用复杂的递归方法,导致算法的收敛性能较差,难以达到高效传输的目的。

为了解决上述问题,本文提出了一种基于热点度轨迹显影机制的网络社区聚类算法。根据网络社区聚类具有的多径一体特性,采用抽样方法来强化对热点信号的捕捉,并结合角度估计方法来实现热点的精确搜寻,提升聚类效率。随后,通过社区网络节点信号发射中形成的传输矩阵,对热点矢量信号空间完成按列重排,并设计了热点度轨迹显影,实现热点的迅速显影,提高聚类的存活时间。最后,

通过 MATLAB 仿真软件,测试了本文算法的有效性。

1 本文算法设计

鉴于当前网络社区聚类算法更多的是基于移动无线传感及数据采集领域^[6],且考虑到当前网络节点正逐步从传统的固定感知网络转向移动感知网络,因此,在本文算法中,信号预发射模型以 LTE-5G 为主^[7],其模型见图 1。在某个具体的时刻,整个社区中有若干个等待聚合的社区节点,需要通过一定的算法进行聚合,且这些节点均需要通过一定的聚合算法进行聚集,以便形成数据互通、热点聚集、传输多径为典型特征的网络社区聚类。此外,由于节点均处于流动状态,如移动无线传感网、移动物联网,二者均采用典型的 LTE-5G 制式网络,因此,在聚合过程中需要考虑链路抖动因素,以便聚类能够正常工作,有效减缓背景噪声的影响。

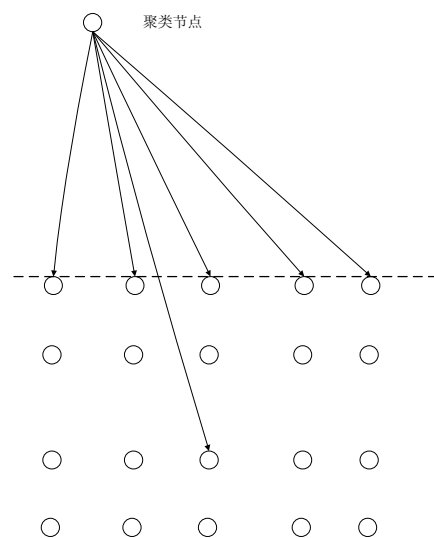


图 1 网络社区聚类

Fig.1 Network community clustering

设 M 个社区节点需要 M 个热点节点来协助,以形成响应聚类,社区节点的接入信号为 $G_1, G_2, \dots, G_M$ 。该信号传输周期为 E_i , 其中 $i=1, 2, 3, \dots, M$ 。则第 M 个社区节点的当前信号特征为:

$$F^{(M)}(\omega) = \sum_{k=1}^{k=m} S(\omega) H^{(k)}(\omega) + H^{(M)}(\omega) \quad (1)$$

其中, $F^{(M)}(\omega)$ 为社区节点预发射中的信道固定频

率; $H^{(M)}(\omega)$ 为社区节点在信道固定频率条件下传输噪声的能量^[8]; $S(\omega)$ 为热点节点的信号能量分布。

根据模型 (1) 热点节点的信号能量分布, 则社区节点当前信号时域特征 $s(t)$ 可由下式获取:

$$s(t) = a_i \sum_{i=1}^{\infty} A(t) \quad (2)$$

其中, a_i 为热点节点对应的最大信号强度值。

根据模型 (2), 基于多径传输理论^[9]可知, M 个社区节点同时获取的热点节点信号为:

$$\begin{aligned} F_1 &= w_1 + E_{\omega_1} S \\ F_2 &= w_2 + E_{\omega_2} S \\ &\dots \\ F_m &= w_m + E_{\omega_m} S \end{aligned} \quad (3)$$

其中, $S = \Lambda[S(\omega_1), S(\omega_2), \dots, S(\omega_\lambda)]^T$ 为强度 λ 的周期抽样, $E_{\omega_i} (i=1, 2, \dots, M)$ 为各热点节点接收到的社区节点能量强度抽样值, 且满足:

$$E_{\omega_1} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 0 & 0 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ e^{-\lambda_j \Delta \omega \omega^{(1)}_1} & 0 & \dots & 0 \end{bmatrix} \quad (4)$$

$$E_{\omega_2} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 0 & 0 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ e^{-\lambda_j \Delta \omega \omega^{(2)}_2} & 0 & \dots & 0 \end{bmatrix} \quad (5)$$

$$E_{\omega_m} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 0 & 0 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ e^{-\lambda_j \Delta \omega \omega^{(m)}_m} & 0 & \dots & 0 \end{bmatrix} \quad (6)$$

设 $\Delta t_{i,j} = t_i - t_j$ 为社区节点搜寻热点节点的时延^[10], 其搜寻过程中的坐标角度模型见图 2, 则 $\Delta t_{i,j}$ 为:

$$\Delta t_{i,j} = d_{i,j} \sin \theta / c \quad (7)$$

其中, $d_{i,j}$ 为社区节点搜寻热点节点时最大搜索距离。

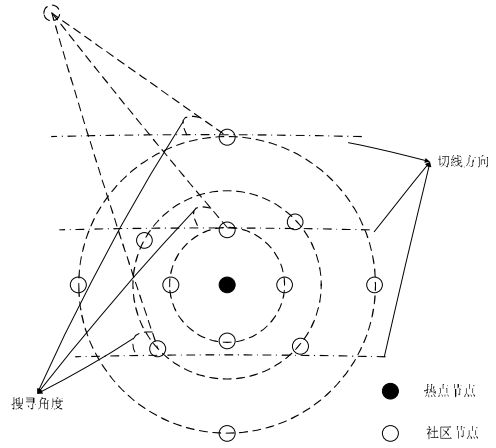


图 2 社区节点搜寻热点节点

Fig.2 Community node search hotspot node

随后, 根据模型 (7) 可得 $\bar{\theta}_{i,j}$:

$$\bar{\theta}_{i,j} = \arcsin(c \Delta t_{i,j} / d) \quad (8)$$

对网络中全部社区节点按模型 (3) 获取的矩阵进行轨迹划分, 划分方式如下:

$$E_{i,j} = \begin{bmatrix} w_i \\ SE_{\omega_i} \end{bmatrix}^T + \begin{bmatrix} w_j \\ SE_{\omega_j} \end{bmatrix}^T \quad (9)$$

令 $D(t_i, t_j) = \begin{bmatrix} SE_{\omega_i} \\ SE_{\omega_j} \end{bmatrix} \in \Omega$, $D(i, j) = \begin{bmatrix} w_i \\ w_j \end{bmatrix} \in \Omega$, 则模型 (9)

变为:

$$D_{i,j} = w(i, j) + D(t_i, t_j) \quad (10)$$

令 $\bar{M} = D_{i,j} D_{i,j}^H$, $D_{i,j}^H$ 为 $D_{i,j}$ 的莱斯变换, $\bar{M}$ 的莱斯映射有 $2M$ 个^[8]. 则对 $\bar{M}$ 按照莱斯映射后进行重排序, 可得显影矩阵:

$$\bar{M} = \bar{M}_s + \bar{N}_s \quad (11)$$

其中, $\bar{M}_s \in \Omega^{2M}$ 为社区节点信号矩阵; $\bar{N}_s \in \Omega$ 为热点聚类. 对于任意社区节点 i , 按照模型 (11) 进行按列重排后, 第 i 个聚类对应的热点节点为社区节点 i 所对应的热点节点, 从而在 M 个热点节点中实现待搜寻热点的显影。

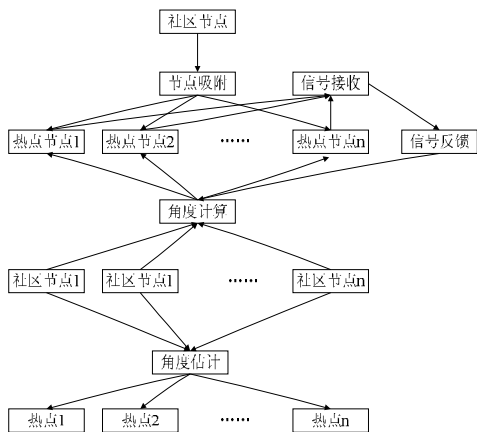


图3 本文算法流程图

Fig.3 The flow chart of algorithm

2 仿真实验

为便于验证本文算法的效果,采用 MATLAB 软件实施测试。为了突出所提算法的优势,将聚类流动性映射算法^[12] (Clustering Liquidity Mapping Algorithms, CLM 算法)、超欧里几何热度聚类算法^[13] (Hyper-Eulerian Geometric Thermal Clustering Algorithms, H-EGTC 算法)作为对照组。实验参数设置如下:网络节点采用随机分布模型,节点密度不低于 $0.1 \text{ 个}/\text{m}^2$,网络通信信号采用 LTE-5G 信号,中心频率为 1.024 GHz ,信道信噪比不低于 16 dB 。评估指标为聚合时间、热点显示时间、搜寻失误差率、能耗。

2.1 聚合时间

图4为聚合时间仿真结果。依图发现,本文算法具有更低的聚合时间,显著低于 CLM 算法及 H-EGTC 算法。这是由于本文算法采用轨迹划分方式直接对各个热点进行显影,可显著提高社区节点匹配热点的效率,降低单个社区节点成功聚合的周期;而且在搜寻过程中采用了角度估计方法来提高聚合精确度,能够适应高流动性的场景,要优于 CLM 算法单纯采用的周期聚方法,以及 H-EGTC 算法采取的拓扑聚合机制。CLM 算法采用周期聚合方式,需要根据聚类生成周期等多个维度来设定聚合阈值,难以适应流动性较高的实际环境。H-EGTC 算法需要在统计节点热度的情况下对拓扑状况进行实时捕捉,在节点运动速度较高时,容易存在信噪比难以控制的缺陷,因此聚合时间要

高于本文算法。

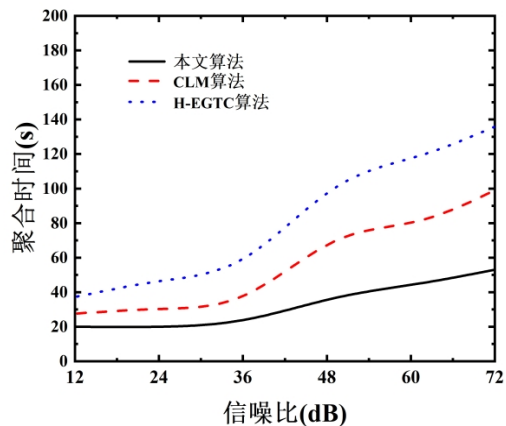


图4 聚合时间测试

Fig. 4 Aggregation time test

2.2 热点显示时间

图5为不同算法的热点显示时间测试结果。由图发现,本文算法热点显示时间要显著高于 CLM 算法及 H-EGTC 算法。原因是本文算法采用的基于矩阵方式的轨迹划分能够显著提高热点捕捉效果,搜寻热点仅需要通过排序机制由高到低进行直接搜寻,有效降低了搜寻过程的复杂程度,因此热点聚类效果更好。此外,本文算法在搜寻过程中能够根据角度估计实现精确定位,捕捉热点的能力较强,可在高强度信道噪声干扰环境下进行热点聚类,具备很强的聚合能力。而 CLM 算法通过预设连通度阈值方式进行热点搜寻,当仅当聚类低于连通度阈值时才能进行热点搜寻,效率较低;H-EGTC 算法主要通过检测拓扑联通方式进行热点搜寻,没有深入到信号发射层次进行热点聚类,容易因信道噪声污染而导致热点显示异常,因此显示性能要低于本文算法。

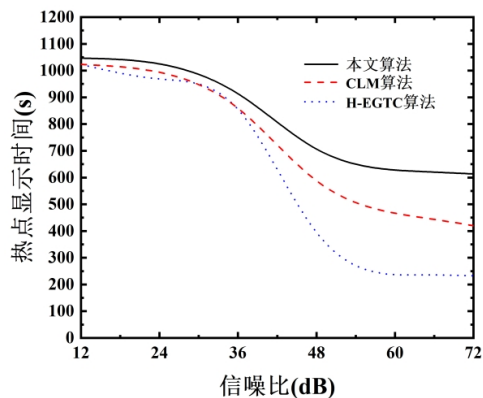


图5 热点显示时间测试

Fig.5 Hotspot display time test

2.3 搜寻失误率

图 5 为不同算法的搜寻失误率仿真结果。根据图中数据发现,本文算法搜寻热点过程中的失误率始终要低于对照组。这是因为本文算法采用轨迹划分方式能够直接对热点进行显影,提高社区节点匹配热点的效率,能在较短时间内对热点的精确聚合。另外,本文算法的轨迹划分方式主要基于矩阵形式,只需要进行简单匹配就能迅速实现对热点的匹配,使其搜寻失误率处理较低的水平,且搜寻过程中热点显示时间及聚合时间也较好,能够在高干扰环境下实现对热点的精确搜寻,降低搜寻失误率。而 CLM 算法搜寻热点是按照初始设定的连通度阈值进行搜寻的,难以适应热点性能的动态变化,使其匹配效果不佳。H-EGTC 算法需要周期搜寻机制进行热点搜寻,难以在搜寻过程中进行抗噪处理,因此搜寻失误率要高于本文算法。

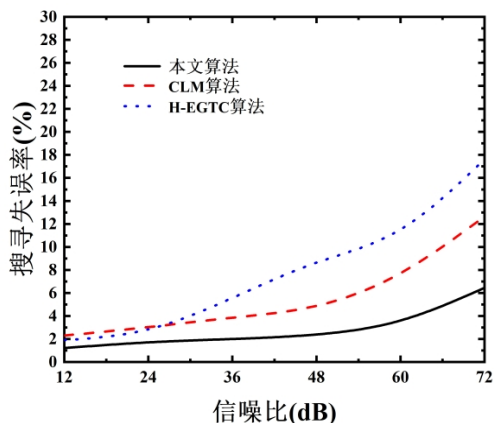


图 6 搜寻失误率测试

Fig.6 Search error rate test

2.4 能耗

图 6 为不同算法的能耗测试结果。依图可知,本文算法能耗最低。原因是本文算法采用的轨迹划分方式除了能够直接对热点进行显影之外,其使用的矩阵分割方式还能够在确保搜寻正确率的情况下迅速实现热点捕捉,具有聚合时间短的特点,因此聚合过程中的能耗较低。CLM 算法采用的阈值初始设置方式无法动态实现热点精确捕捉,因此具有能耗较大的特点;H-EGTC 算法采用周期搜寻机制,其搜寻难度随着热点搜寻失误的增加而增加,难以将能耗控制在较低的水平。因此,两种对照组算法的能耗均要高于本文算法。

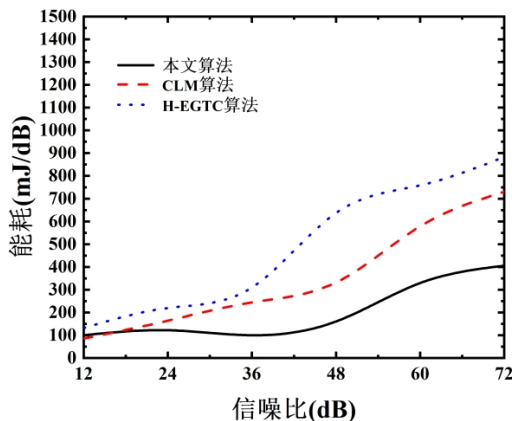


图 7 能耗测试

Fig.7 Energy consumption test

3 结束语

针对当前网络社区聚类算法具有聚合收敛度较低,热点显示困难,能耗水平难以控制等不足,提出了一种基于热点度轨迹显影机制的网络社区聚类算法。主要通过信号抽样方式构建热点信号节点接收矩阵,并使用正交分割方式进行轨迹显影,能够显著提高热点捕捉效率,大大提高聚类生成质量,且具有能耗较低的优势。

下一步,将考虑到本文算法在物联网领域运用的一些困难,如难以适应自组织方式组织基于区块链分叉方式的能耗结算模型,拟采用基于二叉树方式的分叉机制进行连通度建模,提高算法在高强度能耗条件下对物联网环境的适应能力,增强本文算法的实际部署价值。

参考文献:

- [1] 乔少杰,郭俊,韩楠,等. 大规模复杂网络社区并行发现算法[J]. 计算机学报,2017,40(3):687-700.
- [2] 刘浩广,伍家驹. 基于FLD算法的南昌市老城区公交网络的优化研究[J]. 井冈山大学学报:自然科学版,2017,38(6): 55-60.
- [3] Kai L, Zhi G. Image Segmentation with Fuzzy Clustering Based on Generalized Entropy[J]. Journal of Computers, 2014, 9(7): 1678-1683.
- [4] Zhi Z Z, Zhen Y W. Mining Overlapping and Hierarchical Communities in Complex Networks[J]. Physical A: Statistical Mechanics and its Applications,2015,421(41): 296-311.

-
- [5] Chun C L, Jia R K, Jyun Y C. An Integer Programming Approach and Visual Analysis for Detecting Hierarchical Community Structures in Social Networks[J]. Information Sciences, 2015, 299(3): 25-33.
- [6] Yi C, Xiao L W, Xin X, et al. Overlapping Community Detection in Weighted Networks Via a Bayesian Approach[J]. Physical A: Statistical Mechanics and its Applications, 2017, 468(12): 790-801.
- [7] Claudio M R. Effects of Multi-state Links in Network Community Detection[J]. Reliability Engineering and System Safety, 2017, 163(12): 46-56.
- [8] Wei Z, Xiao K Z, Zhao K. Analysis of Associtivity and Community Structure in Mobile Social Networks[J]. Procedia Computer Science, 2017, 107(6): 630-635.
- [9] Yang J L, Huang T, Song W M. Discover the Network Mechanisms Underlying the Connections Between Aging and Age-Related Diseases[J]. Scientific reports, 2016, 23(6): 32-56.
- [10] Guo J, Feng C. Implementation of Envelope Analysis on a Wireless Condition Monitoring System for Bearing Fault Diagnosis[J]. International Journal of Automation and Computing, 2015, 12(01): 14-24.
- [11] Tzy S W, Hui T L, Ping W. Weighted-Spectral Clustering Algorithm for Detecting Community Structures in Complex networks[J]. Artificial Intelligence Review, 2017, 47(4): 463-483.
- [12] Shu Y, Zhang L W, Liang L. Discovering Similar Chinese Characters in Online Handwriting with Deep Convolutional Neural Networks[J]. International Journal on Document Analysis and Recognition (IJDAR), 2016, 19(3): 237-252.
- [13] Ze X Z, Xia J, Richard N. Discovering Causal Interactions Using Bayesian Network Scoring and Information Gain[J]. BMC Bioinformatics, 2016, 17(1): 1021-1036.
-

(上接第 28 页)

- [22] 段玉坤,伍志强,韩为东,等. PKR 诱导细胞凋亡发生机制的研究进展[J]. 现代肿瘤医学, 2011, 19(9): 1863-1865.
- [23] 翟景波,高巍,王秋波,等. RNA 传感器蛋白催化酶 PKR 研究新进展[J]. 微生物学杂志, 2016, 36(6): 93-97.
- [24] Nakamura T, Furuhashi M, Li P, et al. Double-stranded rna-dependent protein kinase links pathogen sensing with stress and metabolic homeostasis[J]. Cell, 2010, 140(3): 0-348.
- [25] 胡有生. DsRBM 对草鱼 PKR 抑制蛋白翻译功能的影响研究[D]. 南昌:南昌大学, 2016.
- [26] Stéphanie Dabo, Eliane F. Meurs. dsRNA-dependent protein kinase pkr and its role in stress, signaling and HCV infection [J]. Viruses, 2012, 4: 2598-2635.
- [27] Hu Y, Fan L, Wu C, et al. Identification and function analysis of the three dsRBMs in the N terminal dsRBD of grass carp (*Ctenopharyngodon idella*) PKR[J]. Fish & Shellfish Immunology, 2016, 50: 91-100.
- [28] 李雯. 草鱼 PKR 基因的克隆、表达特性及功能分析[D]. 南昌:南昌大学, 2013.
- [29] Rothenburg S, Deigendesch N, Dey M, et al. Double-stranded RNA-activated protein kinase PKR of fishes and amphibians: Varying the number of double-stranded RNA binding domains and lineage-specific duplications[J]. Bmc Biology, 2008, 6(1): 1-19.
- [30] 石华,宋方洲. 双链 RNA 依赖性蛋白激酶的结构与作用[J]. 生命的化学, 2006, 26(1): 38-41.
- [31] Garcia-Ortega M B, Lopez G J, Jimenez G, et al. Clinical and therapeutic potential of protein kinase PKR in cancer and metabolism[J]. Expert Reviews in Molecular Medicine, 2017, 19: 9.
- [32] 梅雯,张成桂,自加吉,等. *MTERF1* 基因真核表达载体的构建及其在 C-33A 细胞中的表达研究[J]. 井冈山大学学报:自然科学版, 2017, 38(5): 29-34.