

基于概率分布的复杂网络结构模块化方法研究

刘丙利, 郭雯雯

(郑州财经学院 信息工程学院, 河南 郑州 450000)

摘 要: 提出基于概率分布的复杂网络结构模块化方法. 复杂网络结构模块化拥有一个专属流程, 分为模块变量分配和结构化索引两个步骤. 在模块变量分配中, 所提方法构建出复杂网络模块概率分布模型, 通过吉布斯采样获取模型学习样本. 对样本进行监督, 以获取模块条件概率函数真实解. 用真实解表示模块比重, 真实解的数量表示复杂网络结构模块数. 在结构化索引中, 所提方法利用 LDA 主题模型确定以上构建的模块化复杂网络是否成立. 实验结果表明, 所提方法构建出的模块化复杂网络具备较强线性.

关键词: 概率分布; 复杂网络; 模块化; 吉布斯采样; LDA 主题模型

中图分类号: TP311.13

文献标识码: A

文章编号: 1000-7180(2017)06-0114-04

Research on Modular Method of Complex Network Structure Based on Probability Distribution

LIU Bing-li, GUO Wen-wen

(Information Engineering College, Zhengzhou Institute of Finance and Economics, Zhengzhou 450000, China)

Abstract: The complex network node and a large number of data constantly changing complex network structure of the modular method, previously proposed mostly lack of universality and linear, cannot effectively improve network performance index engine. In this paper, a method of complex network structure based on probability distribution is proposed. The modular structure of complex network has an exclusive process, which is divided into two steps: module variable allocation and structured index. In the module variable assignment, the proposed method constructs a probability distribution model of complex network module, and obtains the model learning sample by Gibbs sampling. The sample is monitored to obtain the real solution of the conditional probability function. The number of real solutions is represented by the number of real solutions. In structured index, the proposed method uses the LDA topic model to determine whether or not the modular complex network is established. The experimental results show that the modular complex network constructed by the method has strong linear.

Key words: probability distribution; complex network; modularization; gibbs sampling; LDA topic model

1 引言

复杂网络中的信息过于繁多, 索引结果中常夹杂一些无效信息, 所以大多数人不能满足现有的信息索引模式^[1]. 在现有状态下, 若想删除复杂网络中的无效信息是不可能的, 只能通过缩小复杂网络结构索引区间改善索引精确度, 即令复杂网络结构模块化^[2].

文献[3]中表述的路径挖掘方法和文献[4]中表

述的理论场模型, 都能将复杂网络结构模块化, 但这些早期的方法往往运算量大, 模块化线性低. 文献[5]提出的基于线性规划的复杂网络结构模块化方法可提升模块化线性, 但存在一定的使用局限. 故需要提出一些新的、有效的模块化方法满足网络用户的索引需求,

2 复杂网络结构模块化

进行复杂网络结构模块化有一个专属流程, 如

图1所示,图中的矩形代表模块.从复杂网络中随机提取一组动态变量 $X = \{X_1, X_2, \dots, X_n\}$, 数量计 n 个,对于其中任意变量 X_i , 它的定义域用 $val(X_i)$ 表示.假设 C 是一个复杂网络结构模块化集合, 归属于集合 C 中的模块共有 k 个, 用 $M_1, M_2, \dots, M_k$ 表示, 将其中第 j 个模块的变量定义域表示为 $val(M_j)$. 处于相同模块中的变量使用同一个条件概率函数、模块配置和母节点, 区别不同变量性质的模块属性是子节点的不同, 比如, X_2, X_3, X_4 的母节点都是 X_1 , X_3 另外拥有子节点 X_6 和 X_7 , 而 X_2, X_3 的子节点却不是 X_6 和 X_7 或无子节点.

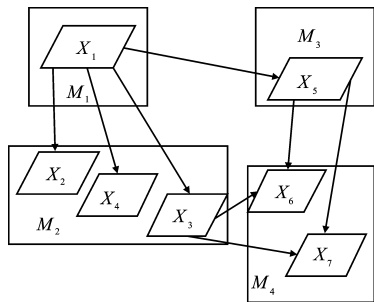


图1 复杂网络结构模块化专属流程图

图1中,集合 C 中包含4个模块, $M_j \in C$, 四个模块使用的条件概率函数 $Pa(M_j)$ 不相同, 但均满足 $Pa(M_j) \subset X$. 如果这时存在 $val(X_i) = val(M_j)$ 的线性关系, 则复杂网络结构的模块化复杂网络是成立的.

为了更通俗易懂地表达复杂网络结构模块化的专属流程, 将图1大致分为两个步骤:

- (1) 将复杂网络结构变量分配到模块内;
- (2) 进行结构化索引, 确定模块化网络是否成立.

3 基于概率分布的复杂网络结构模块化方法研究

3.1 基于概率分布的变量分配方法

概率分布归属于概率论中的一项基本概念, 用来定义某一变量数值的概率情况. 概率分布的表现形式具有多样性, 常见的有随机型、离散型和连续型^[6]. 复杂网络结构变量的概率分布复杂多变, 不能逐字逐句地对其进行表征和分析, 因此需要定义一个联合概率分布 $f(X_i)$ 用来表征变量性质的不确定性, 在 $f(X_i)$ 之上构建复杂网络概率分布模型, 整个过程用图2描述.

由图2可知, 基于概率分布的变量分配方法通

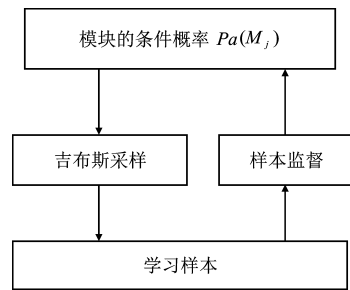


图2 复杂网络模块概率分布模型的构建

过吉布斯采样和机器学习两个过程构建复杂网络模块概率分布模型. 赋予条件概率函数 $Pa(M_j)$ 一个初始值, $Pa(M_j) < f(X_i)^{[7]}$, 使用吉布斯采样法获取模型学习样本, 对样本进行监督. 循环这个过程, 得到 $Pa(M_j)$ 的真实解. $Pa(M_j)$ 真实解就是模块比重, 有几个真实解, 复杂网络结构就将被划成几个模块.

3.1.1 吉布斯采样

吉布斯采样的目的是令模块条件概率无限趋近联合概率分布. 模糊、多次提取某一个复杂网络结构变量坐标, 求取坐标平均值用来表征变量方位:

$$E(X_i) = \lim_{T \rightarrow \infty} \frac{\sum_{t=1}^T \beta(X_t)}{T} \quad (1)$$

设吉布斯采样初值为 N , 其伪代码流程如下:

(1) 将模块条件概率函数 $Pa(M_j)$ 初始值输入, 令 $t = 1 \sim (N + K_0)$, K_0 是采样增值, $\sim$ 表示按位取反;

(2) 循环进行 $P = E(X_i) \cup Pa(M_j | X_i(t))^{[8]}$ 的计算, 等式左边的 P 是学习样本集合, 初值为空. 等式右边的 $Pa(M_j | X_i(t))$ 表示条件概率函数稳态值, $X_i(t)$ 时 t 采集时刻复杂网络结构变量属性即时值.

(3) 当 $t > K_0$, 循环停止, 输出学习样本 P .

3.1.2 样本监督

样本监督是指将学习样本 P 中存在明显属性的复杂网络结构变量引出并进行机器学习, 其过程用式(2)描述:

$$Q_i = \{(x^i, x_i) \mid x \in P\} \quad (2)$$

式(2)实际上是在进行模块条件概率函数 $Pa(M_j)$ 真实值的估计, Q_i 为估计值, x 是 P 中存在明显属性的变量集合, x_i 是 x 中任意一个变量, x^i 是 x_i 的期望. 对 Q_i 进行修正可求解 $Pa(M_j)$ 的真实值, 如果设变量 X_i 取值满足:

$$X_i \in \{-1, 1\} \quad (3)$$

杂网络的主题则与文学有关。

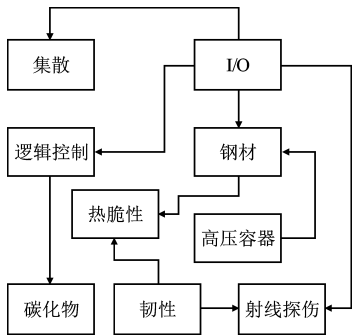


图 4 模块化复杂网络(模块数量为 9 个)

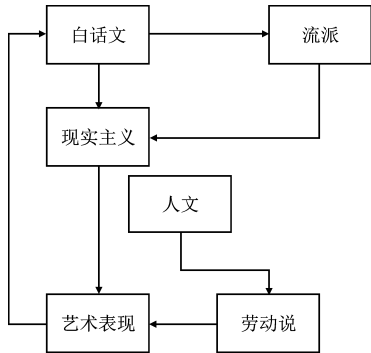


图 5 模块化复杂网络(模块数量为 6 个)

当加快某一个主题的数据行为时,其他主题的数据行为也应该产生线性变化.实验提高了“集散”主题和“人文”主题的数据行为,得到其他主题的变化情况,如表 1、表 2 所示,可以看出,本文方法构建的模块化复杂网络具备较强线性.

表 1 复杂网络结构模块化线性分析结果 1

主题名称	“集散主题”提高倍数/倍			
	0.1	0.5	2	5
I/O	0.435	0.218	3.144	7.862
逻辑控制	0.503	0.252	3.168	7.922
钢材	0.461	0.231	3.843	9.608
热脆性	0.483	0.242	2.088	5.215
高压容器	0.112	0.055	1.511	3.770
碳化物	0.159	0.081	2.003	5.008
韧性	0.287	0.144	2.580	6.450
射线探伤	0.288	0.144	2.250	5.628

5 结束语

本文提出了基于概率分布的复杂网络结构模块化方法,通过变换研究方向,从复杂网络概率分布入手进行研究,采用吉布斯采样和 LDA 主题模型等手段,有效增强了模块化复杂网络的线性.

表 2 复杂网络结构模块化线性分析结果 2

主题名称	“人文”主题提高倍数/倍			
	0.1	0.5	2	5
白话文	2.179	1.111	0.358	0.876
流派	2.063	1.032	0.203	0.505
现实主义	5.258	2.629	0.200	0.500
艺术表现	1.034	0.510	0.598	1.495
劳动说	0.559	0.280	5.031	12.588

参考文献:

- [1] 王文胜,杨洪勇.复杂网络社团挖掘中基于路径的模块性分析方法[J].计算机工程,2016,42(4):173-178.
- [2] 曹黎侠,黄光球,李艳.第三方支付粗糙复杂网络知识发现方法研究[J].计算机科学与探索,2016,10(8):1143-1153.
- [3] He J L, Yan F, Chen D B. A Novel Top-kStrategy for Influence Maximization in Complex Networks with Community Structure [J]. Plos One, 2015, 10(12):e0145283.
- [4] 黄蓝会.基于多子网复合复杂网络的模型研究[J].电子设计工程,2016,24(14):74-76.
- [5] 李静,张洪欣,王小娟,等.确定度分布条件下可变同配系数的算法构造与影响分析[J].物理学报,2016,65(9):168-177.
- [6] Liu A H, Wu C. Research on risk assessment method of disaster chain based on complex network[J]. Xitong Gongcheng Lilun Yu Shijian/system Engineering Theory&Practice, 2015, 35(2):466-472.
- [7] 叶礼邦,王满喜,张玉灵,等.基于多连接方式的无线通信网拓扑模型研究[J].现代电子技术,2016,39(7):1-4.
- [8] 蒋晓娟,张文生.加权网络的在线结构学习算法[J].模式识别与人工智能,2016,29(2):122-130.
- [9] Tongfei L, Wei L, Jianfeng W, et al. Research on the Characteristic of Automotive Failure Diagnosis Based on Complex Networks[J]. Open Mechanical Engineering Journal, 2015, 9(1):508-513.
- [10] 岳增慧,许海云,方曙.基于微分动力学的科研合作网络知识扩散模型及影响机制研究[J].情报学报,2015,34(11):1132-1142.
- [11] 李婵婵,蒋国平.社团结构网络环境下 SIS 病毒传播建模与分析[J].复杂系统与复杂性科学,2016,13(2):67-73.

作者简介:

刘丙利 女,(1982-),硕士,讲师.研究方向为计算机网络技术及相关应用. E-mail:13526583367@139.com.

郭雯雯 女,(1984-),讲师.研究方向为软件开发.