

杨秋格, 吴鹏, 孙旭光, 等. 基于时间序列的灾害区域关联网络构建方法[J]. 灾害学, 2015, 30(4): 21-24. [Yang Qiuge, Wu Peng, Sun Xuguang, et al. Complex Network Construction Method of Disaster Regional Association based on Time Series[J]. Journal of Catastrophology, 2015, 30(4): 21-24.]

基于时间序列的灾害区域关联网络构建方法*

杨秋格¹, 吴鹏², 孙旭光¹, 陈丹琪¹

(1. 防灾科技学院 灾害信息工程系, 河北 廊坊 065201; 2. 防灾科技学院 信息化管理中心, 河北 廊坊 065201)

摘要: 针对灾害地区相关性问题, 提出一种灾害区域关联复杂网络构建方法。首先利用幂级数展开得出网络灾害节点的系统动力学方程, 通过节点关联系数来表征节点间的区域关联关系, 并通过相关转化使得可以利用压缩感知理论对节点关联系数进行求解, 从而实现网络节点关系拓扑的构建。最后通过中国地震灾害数据进行性能仿真实验, 实验结果表明, 完成网络构建只需要少量的时序信息, 构建得出的区域关联结果也具有一定的科学性。

关键词: 灾害; 时间序列; 压缩感知; 区域关联; 复杂网络

中图分类号: X43; TP18 **文献标志码:** A **文章编号:** 1000-811X(2015)04-0021-04

doi: 10.3969/j.issn.1000-811X.2015.04.004

自然灾害的区域关联现象一直是灾害学研究的热点, 大量观察到的事实, 如历史记录中某些灾害(如地震)在部分地区会表现出一种同步衰涨趋势, 即同时趋于活跃或平静^[1], 以及2005年及2007年的苏门答腊地震后, 我国南方都随之发生了雨涝灾害, 这表明了旱、洪、震等自然灾害存在着一定的区域关联性^[2], 而对其的研究则能够为灾害预测提供支持。

灾害的时间序列是进行灾害区域关联研究的基本信息, 通过对两个地区的灾害时间序列进行相似性匹配, 从而确定两者之间是否具有相关性, 也是近年来灾害区域关联分析的常见方法^[3-5]。由于这类方法每次只选取两个区域进行相似性匹配(即单点对单点), 因而在时间序列信息上会有大量的截取操作, 而如果我们想以复杂网络角度来研究多个区域间的灾害相关性(即多点对多点), 由于多个点间同步影响会形成一个复杂系统, 就使得这种单点匹配得到的区域相关性难以作为节点链接来构建灾害区域关联复杂网络。

如何基于各区域的灾害时序信息, 挖掘出潜在的关联关系并重构网络, 是本文所要解决的问题。事实上, 复杂网络重构正逐渐成为复杂网络研究的热点问题, 已提出的重构方法如逆向工程法^[6]、微扰响应法^[7]、收敛镜像系统法^[8], 以

及基于噪声的相关性方法等^[9], 其基本思想都是通过个体单元行为来推导网络拓扑, 但都或多或少存在着运算成本过大的问题。文献[10-12]中则针对这一问题, 提出一种基于压缩感知的复杂网络重构方法, 能够以相对于网络规模很少量的数据实现拓扑重构, 并对非线性系统及博弈网络等进行了重构尝试, 取得不错的效果。

本文则借鉴压缩感知思想, 基于灾害的时间序列信息, 将灾害区域视为复杂网络节点, 通过构建出节点的复杂系统动力学方程, 进而确定节点间的拓扑关系, 从而构建出一个灾害区域关联复杂网络。在后面的实验中我们会看到, 只需要少量的时序信息, 本文方法就可以完成对网络的构建工作。

1 压缩感知

作为一种新的信息获取理论, 压缩感知技术主要基于可压缩信号的稀疏性, 并满足在远小于Nyquist采样率的情况下, 实现对信号的精确重建。压缩感知技术的主要优势在于能够只基于少量观测数据 $S(S \in R^M)$ 就实现对稀疏向量 $a(a \in R^N)$ 的

* 收稿日期: 2015-04-13 修回日期: 2015-05-21

基金项目: 中央高校青年教师资助计划项目(ZY20130213); 中央高校创新团队资助计划项目(ZY20120104)

作者简介: 杨秋格(1981-), 女, 山东聊城人, 硕士, 讲师, 研究方向为物联网技术. E-mail: yangqiuge0302@163.com

高效恢复,并满足 $S = D \cdot a$,其中 D 是一个 $M * N$ 维满秩矩阵,而信号重建过程即为对以下凸优化问题进行求解,如式(1)所示。

$$\min \|a\|_1 \quad \text{subject to } D \cdot a = s, \quad (1)$$

式中: $\|a\|_1 = \sum_{i=1}^N a_i$ 是向量 a 的 L_1 范数。

由于当 a 为稀疏向量并使得 D 满足式(1)时,利用压缩感知技术可以在 M 远小于 N 的情况下,准确恢复出 a 。而鉴于复杂网络本身表现出的稀疏性,保证了向量 a 的稀疏性,那么只要把复杂网络重构问题转化成压缩感知理论能够处理的形式,就可以利用这一方法基于少量的时序信息准确重构网络拓扑。

对于灾害区域关联网,由于各个区域节点的灾害信息是可以获取的,而具体灾害结果的造成除了受区域自身各种因素的影响,还类似于振子网络,同时受到其他多个区域多种因素的复合作用。因此本文希望能够用少量的相关时序信息,从多节点复合影响的角度,推演出各个节点间的关联情况,进而构建出整个网络拓扑。

2 基于时序的网络构建算法

设灾害区域关联复杂网络中节点个数为 n ,第 i 个节点在某一时刻(如时刻 t)的时序信息被划分为两部分,第一部分设为 s_i ,表示在时刻 t 该区域观测到的受灾结果,第二部分设为 x_i ,这是一个观测参数向量,包含了在时刻 t 与该区域相关的孕灾环境参数分量,如气压、气温、降水、位移场、地应力场等,设该向量的维数为 m 。那么节点 i 的灾害复杂系统动力学方程可以用式(2)表示。

$$S_i = f_i(x_i) + \sum_{j=1, j \neq i}^n G_{ij}(x_j - x_i), \quad (2)$$

式中: $f_i(x_i)$ 为节点 i 独自的灾害动力学方程,式(2)后半部分则为节点 i 与其他节点间的关系动力学方程,两者共同组成节点 i 的系统动力学方程; G_{ij} 表示节点 i 和节点 j 间的观测参数向量关联矩阵,如式(3)所示。

$$G_{ij} = \begin{bmatrix} g_{ij}^{1,1} & g_{ij}^{1,2} & \cdots & g_{ij}^{1,m} \\ g_{ij}^{2,1} & g_{ij}^{2,2} & \cdots & g_{ij}^{2,m} \\ \cdots & \cdots & \cdots & \cdots \\ g_{ij}^{u,1} & g_{ij}^{u,2} & \cdots & g_{ij}^{u,m} \end{bmatrix}, \quad (3)$$

式中: $g_{ij}^{u,w}$ 表示节点 i 观测向量的第 u 个分量与节点 j 观测向量的第 w 个分量间的关联性。如果矩阵中至少存在着一个不为 0 的元素,那么就可以认为这两个节点相互关联,即在复杂网络中存在连边,反之则不存在关联。因此本文的主要工作之一就是求出该关联矩阵是否存在非零值。

对于式(2),我们还可以用式(4)的方式表示

$$s_i = [f_i(x_i) - \sum_{j=1, j \neq i}^n G_{ij}(x_i)] + \sum_{j=1, j \neq i}^n G_{ij}x_j, \quad (4)$$

那么式(4)的前半部分就变为只关于 x_i 的方程,后半部分则是关于其他节点观测向量的方程。此时我们可以将式(4)的前半部分用 $Q_i(x_i)$ 表示,并使用一个 N 阶幂级数将 $Q_i(x_i)$ 展开,如式(5)所示。

$$\begin{aligned} Q_i(x_i) &= f_i(x_i) - \sum_{j=1, j \neq i}^n G_{ij}(x_i) \\ &= \sum_{l_1=0}^N \sum_{l_2=0}^N \cdots \sum_{l_m=0}^N [(a_i)_k]_{1, l_1, l_2, \dots, l_m} \cdot \\ &\quad [(x_i)_1]^{l_1} [(x_i)_2]^{l_2} \cdots [(x_i)_m]^{l_m}, \end{aligned} \quad (5)$$

式中: $(x_i)_1$ 表示节点 i 的观测向量的第 1 个分量, $(a_i)_k$ 表示第 k 个展开项的系数,需要注意的是式(5)涵盖了该阶幂级数下所有可能的展开项,即展开项数目为 $(1 + N)^m$,这意味着许多展开项前的系数可能为 0。

此时,式(4)又可以表示如式(6)所示。

$$S_i = Q_i(x_i) + G_{i1}x_1 + G_{i2}x_2 + \cdots + G_{in}x_n, \quad (6)$$

从式(6)可以看到,在节点时序信息已知的情况下,未知的即为 $Q_i(x_i)$ 中的展开项系数以及关联矩阵 G_{ij} ,同时,为了尽可能地从少量时序信息中构建出网络拓扑,依据压缩感知理论的稀疏性需求,即满足大部分系数为零的条件,我们把式(6)中的 $G_{ij}x_j$ 同样按照式(5)的方式展开,那么式(6)又可以表示如式(7)所示。

$$S_i = Q_1(x_1) + Q_2(x_2) + Q_3(x_3) + \cdots + Q_n(x_n), \quad (7)$$

式(7)的好处在于,它不仅满足了压缩感知理论对稀疏性的需求,同时体现出了节点 i 与其他节点的关联性,如果求出 $Q_i(x_j)$ 的系数全为零,那么就表示节点 i 与节点 j 没有区域关联性,反之如果存在不为零的系数,就表示两者具有区域关联性,同时还能通过分析非零系数所对应的观察向量的分量,进一步研究各个参数对灾害区域关联的影响力。

对于式(7)中各个 $Q(x)$ 函数,如以求解 $Q_i(x_i)$ 为例,即式(5)。设展开项的各个系数用向量 a_i 表示,由于是按照幂级数展开,因此展开形式是固定的,如在级数 $N = 3, m = 3$ 的情况下, $Q_i(x_i)$ 的展开如式(8)所示。

$$\begin{aligned} Q_i(x_i) &= (a_i)_{000} b_i^0 y_i^0 z_i^0 + (a_i)_{001} b_i^0 y_i^0 z_i^1 + \cdots + (a_i)_{uvw} b_i^u y_i^v z_i^w + \cdots \\ &\quad + (a_i)_{333} b_i^3 y_i^3 z_i^3, \end{aligned} \quad (8)$$

式中: b, y, z 为向量 x_i 的 3 个分量。系数向量 a_i 即为 $a_i = [(a_i)_{000}, (a_i)_{001}, \dots, (a_i)_{333}]^T$,同时,对于某一时刻 t 下的观测向量 $x_i(t)$,我们可以得到式(9)。

$$d_i(t) = \left[b_i(t)^0 y_i(t)^0 z_i(t)^0, b_i(t)^0 y_i(t)^0 z_i(t)^1, \dots, \right]. \quad (9)$$

此时 $Q_i[x_i(t)] = d_i(t) \cdot a_i$,如果我们记录有关于节点 i 的 L 个时刻的时间序列信息,那么就可以设观测结果向量为 S ,即 $S = [s_i(t_1), s_i(t_2), \dots, s_i(t_L)]^T$,从而可以得出符合压缩感知理论的 $S =$

$D \cdot a$ 形式的等式,如式(10)所示。

$$\begin{bmatrix} s_i(t_1) \\ s_i(t_2) \\ \dots \\ s_i(t_L) \end{bmatrix} = \begin{bmatrix} d_1(t_1) & d_2(t_1) & \dots & d_n(t_1) \\ d_1(t_2) & d_2(t_2) & \dots & d_n(t_2) \\ \dots & \dots & \dots & \dots \\ d_1(t_L) & d_2(t_L) & \dots & d_n(t_L) \end{bmatrix} \times \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{bmatrix} \quad (10)$$

式中: S 及 $d_i(t)$ 都能够通过已有的时序信息求得,同时由于满足了压缩感知的稀疏性,使得算法可以在时序信息量 L 远小于网络节点规模 n 的情况下求出各系数向量 a ,进而构建出网络拓扑。

3 仿真实验

由于灾害预测研究自身的特点,使得对灾害区域关联网络的有效性验证缺乏绝对参照。针对这一问题,我们选择用具带状分布特点的地震灾害来评估本文方法的性能。我国地域内共计有 23 条地震带,根据这些地震带的分布又可以划分为以下五个主要地震地区:台湾省及其附近海域;西南地区;西北地区;华北地区;东南沿海的广东、福建等地。本文从这 5 个地区中选取共计 60 个城市作为网络节点(除华北地区由于城市密集选取 20 个外,其他地区皆为各选取 10 个城市),时序数据资料来源于历史灾害数据库,包括从 1965 年到 2004 年共计 40 年各个城市的地震受灾情况及相关孕灾环境参数,即 $n = 60, L = 40$ 。

实验平台为 Matlab 7.0。在参数设置上,幂级数阶数 $N = 3$,并根据专家建议选择了 6 项参数作为地震关联参数,即观测向量的维数 $m = 6$,观测结果则根据当年有无 4 级以上地震(含余震)记为 1 或 0。由于地质特点,同一地震带上的节点本身就具备一定的关联性,即每一个地震带可以认为是一个社团,其内部节点具有较高的连接率。根据这一特性,可以从一定程度上检验出本文算法所建立网络的科学性。

图 1 为根据算法运行结果并利用 pajek 软件做出的网络拓扑图,网络的平均度为 3.20,图中出现了 6 个孤立节点,度最大节点为 14 号节点,度为 8,对应城市为中国台北市。

表 1 所示为根据节点关系拓扑,将节点链接映射到节点所对应地震带的连接率统计结果,若本文算法有效,那么首先属于同一地震带的节点间会有较高的连接率。连接率的计算方法为地震带内节点的实际链接数除以节点全连接数。

表 1 各地震带节点连接率

地震带包含节点	地震带数量	平均连接率/%
5	1	60
4	4	75
3	8	72
2	6	83
1	3	0

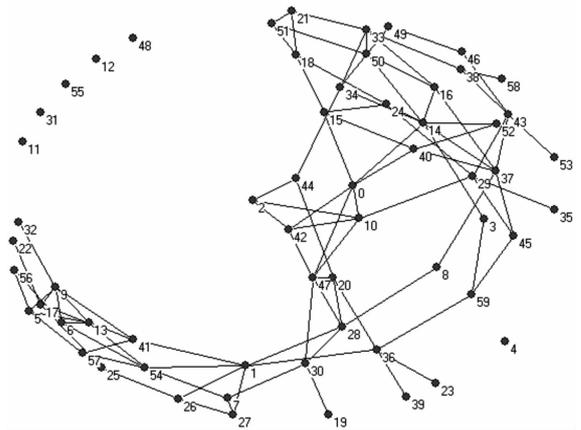


图 1 地震区域关联网络拓扑图

从表 1 可以看到,本文算法建立的网络拓扑中,地震带内节点的连接率是符合期望的,这说明图 1 中的网络拓扑具备了一定的科学性,而对于地震带间出现的节点链接,由于缺乏有效验证方法,因而更适合作为线索信息来对灾害区域关联网络进行分析。这里,通过对 pajek 拓扑图的节点对应区域进行分析,我们发现台湾-四川,四川-云南,青海-西藏这些区域内的节点连接十分密集,其中台湾和四川分属不同的地震区,两者间表现出的关联性值得进一步关注。

为了进一步确认本文算法对时序信息量的低需求。我们用根据初始时序数据(即 $L = 40$)构建的网络拓扑(图 1)作为参照,逐渐减少算法使用的时序信息量,观察算法所建立的网络是否能够和图 1 拓扑保持一致。为了不失一般性,算法分别在每种数据比例下进行 5 次网络构建,时序信息为随机抽取,最后统计与图 1 拓扑匹配度的平均值。

图 2 和图 3 中给出了在不同时序信息量下,所求得网络与图 1 网络的连接匹配度。图中横坐标表示所用数据占原数据量的比例,我们用了两种参数进行评估,分别为节点间的非空链接和空链接。图 2 中纵坐标表示错误链接占所求得链接的比率,图 3 中纵坐标表示求得正确链接占实际链接的比率。

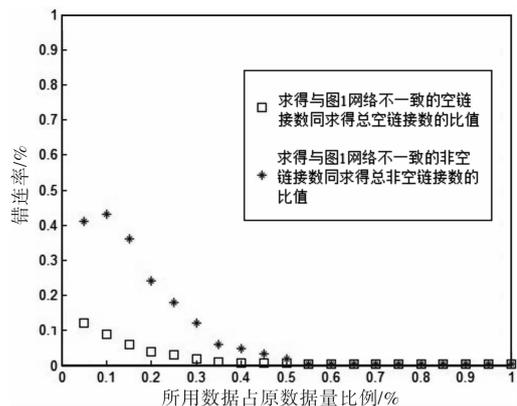


图 2 不同数据量下的错连率

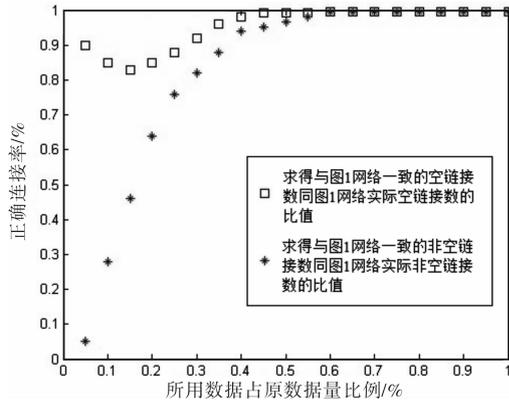


图3 不同数据量下的正确连接率

从图2和图3中可以看到,算法只需要使用60%左右的原时间序列数据量,就可以得出与图1一致的网络拓扑,即实现对图1网络的准确构建。这同时也表明,在保证一定数据量的情况下,算法得出的网络拓扑结果是比较稳定的。

4 结论

本文基于压缩感知理论,尝试利用少量时间序列信息实现对灾害区域关联复杂网络的构建,并取得了一定效果。可能由于所构建网络节点较少的原因,本文实验中所需的数据量尚未达到远小于网络规模的程度。下一步工作将对求得节点动力学方程中的各系数及非零项进行分析,以期能进一步探讨灾害区域关联网络的演化问题。

Complex Network Construction Method of Disaster Regional Association based on Time Series

Yang Qiuge¹, Wu Peng², Sun Xuguang¹ and Chen Danqi¹

(1. Institute of Disaster Prevention, Disaster Information Engineering Department, Langfang 065201, China;
2. Institute of Disaster Prevention, Information Management Center, Langfang 065201, China)

Abstract: Aiming at the disaster regional-related issues, a complex network construction method of disaster regional association based on time series is proposed. The disaster system dynamic equations of network node are obtained through the use of power series expansion and the correlation coefficients between nodes are obtained through the use of compressed sensing theory, so as to realize the construction of the network topology. Experimental results show that, complete network construction requires less amount of time series information and the construction result has a certain rationality.

Key words: disaster; time series; compressive sensing; regional association; complex network

参考文献:

- [1] 门可佩. 重大地震灾害链的时空有序性及其预测研究[J]. 地球物理学进展, 2008, 22(2): 645-651.
- [2] 延军平, 白晶, 苏坤慧, 等. 对称性与部分重大自然灾害趋势研究[J]. 地理研究, 2011, 30(7): 1159-1168.
- [3] 邱剑锋, 谢娟, 李炜, 等. 中强地震的相关性与周期性研究[J]. 计算机工程, 2011, 37(10): 16-18.
- [4] 吴明辉, 许爱强, 周小程, 等. 基于时间序列分析的动调陀螺仪故障预测研究[J]. 计算机测量与控制, 2014, 22(2): 321-324.
- [5] 吴绍春, 吴耿锋, 王炜, 等. 寻找地震相关地区的时间序列相似性匹配算法[J]. 软件学报, 2006, 17(2): 185-192.
- [6] 盛光磊, 甄姬娜. 一种网络信道拓扑结构坚韧度的测量方法研究[J]. 科技通报, 2014, 29(02): 152-154.
- [7] Gardner A T S, d Bernardo D, Lorenz D, et al. Inferring genetic networks and identifying compound mode of action via expression profiling[J]. Science, 2003, 301(5629): 102-105.
- [8] Bongard J, Lipson H. Automated reverse engineering of nonlinear dynamical systems[J]. Proc Natl Acad Sci USA, 2007, 104(24): 9943-9948.
- [9] Timme M. Revealing network connectivity from response dynamics[J]. Phys Rev Lett, 2007, 98(22): 224101-1-4.
- [10] Yu D, Righero M, Kocarev L. Estimating topology of networks[J]. Phys Rev Lett, 2006, 97(18): 188701-1-4.
- [11] Wang W X, Yang R, Lai Y C, et al. Predicting catastrophes in nonlinear dynamical systems by compressive sensing[J]. Phys Rev Lett, 2011, 106(15): 154101-1-4.
- [12] Wang W X, Yang R, Lai Y C, et al. Time-series based prediction of complex oscillator networks via compressive sensing[J]. EPL, 2011, 94(4): 48006-1-6.