

元分析技术及其在软件工程中的应用

卢红敏^{1,2} 许 蕾² 徐宝文²

(东南大学计算机科学与工程学院 南京 210018)¹ (南京大学计算机科学与技术系 南京 210046)²

摘 要 从单个实验研究中很难得到可靠的实验结论,因此需要统计的方法将多个实验研究中的结果进行合并。首先,介绍将多个实验结果进行合并的统计元分析技术。然后,阐述元分析在软件工程研究中的重要意义以及当前的应用现状。最后,分析并总结其中存在的问题。

关键词 元分析,随机效应模型,固定效应模型,实证软件工程

中图法分类号 TP311 文献标识码 A

Meta-analysis Technique and its Application in Software Engineering

LU Hong-min^{1,2} XU Lei² XU Bao-wen²

(School of Computer Science and Engineering, Southeast University, Nanjing 210018, China)¹

(Department of Computer Science and Technology, Nanjing University, Nanjing 210046, China)²

Abstract It is difficult to draw reliable results from a single experimental study. Therefore, there is a need to use a method to combine the results from multiple experimental studies. This paper introduced a meta-analysis technique for this purpose, described its application in empirical software engineering, and pointed out the problems that need to be addressed in the future.

Keywords Meta-analysis, Random-effect model, Fixed-effect model, Empirical software engineering

1 前言

为了从一系列实验研究中得到关于某一特定现象的知识,应对同一现象所做的多个实验结果进行合并,以得到通用化的结论^[1]。元分析是解决该问题的一个理想方案^[1],用来将多个“研究”定量的结果进行合并。在进行元分析时,首先将每个“研究”的结果用同一类型的效应值表示,然后将同类“研究”的效应值以统计的方法进行合并得到平均的效应值。元分析有两个作用:(1)当对同一个现象的多个实验研究得出的结论不一致时,解决这种不确定性;(2)提高对单个实验研究所得结论的信心。本文首先介绍元分析技术,然后阐述元分析方法在软件工程中的应用现状,最后总结其中存在的问题。

2 元分析技术

“固定效应模型”和“随机效应模型”是两种主要的元分析模型,前者假定用于元分析的“研究”是同质的,而后者假定这些“研究”是异质的^[1]。在元分析模型中,单个“研究”的结果称为“可观察效应值”,它是“真实效应值”的一个估算量。固定效应模型的元分析结论只适用于参与元分析的研究,随机效应模型的元分析结论可以推广到没有参与元分析的其他同类研究上。

2.1 固定效应模型

固定效应模型只考虑“研究内方差”。这种模型假定影响

效应值的所有因素在各“研究”中都是相同的,因此这些“研究”具有相同的“真实效应值”。各研究具有相同的“真实效应值”,它们的“可观察效应值”之间的差异仅仅来源于每个“研究”内的随机误差,也即抽样误差。

设 k 是元分析中“研究”的数目, θ 是这些“研究”的“真实效应值”。对第 $i(1 \leq i \leq k)$ 个“研究”,假设 Y_i 和 ϵ_i 各自表示它的“可观察效应值”以及“可观察效应值”与“真实效应值”的偏差。那么,可得:

$$Y_i = \theta + \epsilon_i \tag{1}$$

式中, $E(\epsilon_i) = 0$ 且 $\text{Var}(\epsilon_i) = \sigma_i^2$ 。此处, $E(\cdot)$ 代表一个变量的期望值, $\text{Var}(\cdot)$ 代表一个变量的方差, σ_i^2 是第 i 个“研究”的“研究内方差”。由此,可知 Y_i 的方差为:

$$\text{Var}(Y_i) = \text{Var}(\epsilon_i) = \sigma_i^2 \tag{2}$$

注意,上述公式中 $\text{Var}(Y_i)$ 和 σ_i^2 是种群值,在实践中要用样本值来代替。具体而言,当用样本值 s_i^2 替换式(2)中的 σ_i^2 时,可得到 $\text{Var}(Y_i)$ 的样本估算值 V_i :

$$V_i = s_i^2 \tag{3}$$

为计算 k 个“研究”的“平均效应值” M ,固定效应模型采用方差倒数加权的方式:

$$M = \frac{\sum_{i=1}^k W_i^* Y_i}{\sum_{i=1}^k W_i^*} \tag{4}$$

式中, $W_i^* = 1/V_i$ 为第 $i(1 \leq i \leq k)$ 个“研究”的权值。特别地, M 的方差 V_M 按如下方式计算:

到稿日期:2012-01-20 返修日期:2012-07-20 本文受国家自然科学基金(61073029,90818027)资助。

卢红敏(1975—),女,博士生,主要研究方向为实证软件工程, E-mail: hmlu@seu.edu.cn; 许蕾(1978—),女,副教授,主要研究方向为软件测试; 徐宝文(1961—),男,博士,教授,博士生导师,主要研究方向为软件工程与程序设计语言。

$$V_M = 1 / \sum_{i=1}^k W_i^* \quad (5)$$

2.2 随机效应模型

“随机效应模型”同时考虑“研究内方差”和“研究间方差”。在随机效应模型中,不同“研究”中的“真实效应值”被假定为是不同的并且服从一个正态分布(下文称它的平均值为“总平均值”)。因此,对任一个“研究”而言,其“可观察效应值”与“总平均值”之间的偏差由两部分构成:“真实效应值”与“总平均值”间的偏差(即“研究间方差”)和“可观察效应值”与“真实效应值”间的偏差(即“研究内方差”)。前者刻画各“研究”间“真实效应值”的差异,后者刻画每个“研究”内部的样本误差。元分析的目标是利用统计方法综合各“研究”的“可观察效应值”,得到一个“平均效应值”,用来估算“总平均值”。

设 k 是元分析中“研究”的数目, μ 是这些“研究”的“真实效应值”的“总平均值”。对第 $i(1 \leq i \leq k)$ 个“研究”,假设 Y_i^* 、 ζ_i 和 ϵ_i 各自表示它的“可观察效应值”、“真实效应值”与“总平均值” μ 的偏差以及“可观察效应值”与“真实效应值”的偏差。那么,可得:

$$Y_i^* = \mu + \zeta_i + \epsilon_i \quad (6)$$

式中, $E(\epsilon_i) = 0$ 、 $\text{Var}(\epsilon_i) = \sigma_i^2$ 、 $E(\zeta_i) = 0$ 且 $\text{Var}(\zeta_i) = \tau^2$ 。此处, σ_i^2 是第 i 个“研究”的“研究内方差”, τ^2 是“研究间方差”。由此,可知 Y_i^* 的方差为:

$$\text{Var}(Y_i^*) = \text{Var}(\epsilon_i) + \text{Var}(\zeta_i) = \sigma_i^2 + \tau^2 \quad (7)$$

在实践中,要用样本值 s_i^2 和 T^2 各自替换式(7)中的 σ_i^2 和 τ^2 。于是, $\text{Var}(Y_i^*)$ 的样本估算值 V_i^* 为:

$$V_i^* = s_i^2 + T^2 \quad (8)$$

式中,

$$T^2 = \max\{0, (Q - df) / C\} \quad (9)$$

df 、 Q 和 C 各自表示自由度、总方差和比例因子:

$$df = k - 1 \quad (10)$$

$$Q = \sum_{i=1}^k W_i^* Y_i^{*2} - (\sum_{i=1}^k W_i^* Y_i^*)^2 / \sum_{i=1}^k W_i^* \quad (11)$$

$$C = \sum_{i=1}^k W_i^* - \sum_{i=1}^k W_i^{*2} / \sum_{i=1}^k W_i^* \quad (12)$$

式中, $W_i^* = 1/s_i^2$ 。

随机效应模型采用方差倒数加权的方式计算 k 个“研究”的“平均效应值” M^* 及其方差 V_M^* :

$$M^* = \sum_{i=1}^k W_i^* Y_i^* / \sum_{i=1}^k W_i^* \quad (13)$$

$$V_M^* = 1 / \sum_{i=1}^k W_i^* \quad (14)$$

式中, $W_i^* = 1/V_i^*$ 为第 $i(1 \leq i \leq k)$ 个“研究”的权值。

在元分析中,“研究间方差”称作异质性。 Q 可用来检测各“研究”的“真实效应值”间是否存在异质性。在不存在异质性的零假设下, Q 在统计上服从一个自由度为 $k-1$ 的中心卡方分布。如果 Q 对应的 p 值小于 0.05,那么拒绝零假设,否则不能拒绝零假设。

3 软件工程中的元分析

在软件工程中,实验研究的目的是调查独立变量(例如所使用的方法、语言和工具)和依赖变量(例如质量和效率)间的因果关系。“效应值”指独立变量和依赖变量之间关系的强度,它在样本数据上进行计算,用来描述所调查的现象在种群中出现的程度。在软件工程的实验研究中,典型的效应值包

括相关系数、比值比和平均值差等。元分析常由 3 步构成:选择参与元分析的研究;估算每个研究上的效应值;将各研究上的效应值进行合并得到总体上的效应值。

3.1 意义与价值

在 1999 年,Briand 等将元分析列为软件工程实验研究的四大未来方向之一^[1]。在 2004 年,鉴于软件工程领域内已出现了大量的实验研究,Mair 等指出下一步的工作重点是采用元分析技术将这些实验研究中获得的实验证据进行合并,以获得关于所调查的现象背后的知识^[3]。在 2007 年,Sjöberg 等分析和展望实验方法在 2020—2025 年间对软件工程领域所起的重要作用时,特别强调元分析不仅是实验方法的重要形式,而且是“基于证据的软件工程”中关键的组成部分^[4]。由此可见,元分析在软件工程实验研究中占有越来越重要的地位。

3.2 元分析应用现状

近年来,在软件工程中的元分析研究工作正在明显增多,尽管它们所覆盖的软件工程主题仍然有限。

3.2.1 系统规模与开发工作量的相关性

在 1998 年,Pickard 等对 8 个实验研究进行元分析,调查项目开发工作量和规模之间的相关性^[5]。其中,每一个实验研究的“可观察效应值”是开发工作量(人月)和规模(代码行数)的皮尔森相关系数。他们首先采用固定效应模型,这些研究上的“平均效应值”为 0.83,其 95% 的置信区间为 [0.78, 0.87],该结果说明项目开发工作量和规模间存在强相关。然而,总方差 Q 的 p 值小于 0.05,说明这些实验研究存在“异质性”,因此需要采用随机效应模型进行分析。随机效应模型的分析结果表明,项目开发工作量和规模的“平均效应值”0.81,与固定效应模型所得的结果非常相似。上述结果说明项目开发工作量和规模之间的相关性很高,而且数据来源的异质性对该结论的影响不大。

3.2.2 软件需求规格说明审查方法的有效性

在 1999 年,Hayes 等对 5 个实验研究进行元分析,调查采用问题清单方法(Checklist)和基于场景的方法审查软件需求规格说明的有效性^[6]。在这 5 个实验研究中,先前的研究人员对下述问题进行了调查:(1) 采用问题清单方法和基于场景方法审查软件需求规格说明时,缺陷检测率是否有显著差别?(2) 在两个不同的软件需求规格说明 WLMS 和 Cruise 上,缺陷检测率是否有显著差别?对于前一个问题,有 2 个实验研究报告有显著差别,其余 3 个实验研究报告没有显著差别。对于后一个问题,有 4 个实验研究报告有显著差别,有 1 个实验研究报告没有显著差别。为得到通用的结论,Hayes 等利用固定效应模型对上述两个问题进行元分析。其中,每一个实验研究上的“可观察效应值”为标准化的平均值差。固定效应模型的元分析结果显示,对于审查方法,这些研究的“平均效应值”不显著(95% 的置信区间为 [(0.44, 2.66)]),即基于场景的审查方法和基于问题清单的审查方法在缺陷检测率上没有显著差别;对于规格说明类型,这些研究的“平均效应值”显著(95% 的置信区间为 [0.69, 1.34]),即在不同类型的规格说明上缺陷检测率有显著差别。 Q 检验的结果表明,审查方法元分析的研究间存在“异质性”,规格说明类型元分析的研究间不存在“异质性”。

3.2.3 面向对象度量的相关性及其值

在 2000 年,Succi 等对 100 个开源的 Java 系统进行元分

析,调查 LOC 与圈复杂度度量 $V(G)$ 、LOC 与加权方法数 WMC、DIT 与 Fanout 之间的相关性^[7]。其中,每个系统上的“可观察效应值”为皮尔森相关系数。固定效应模型元分析结果表明,LOC 与 $V(G)$ 、LOC 与 WMC、DIT 与 Fanout 的“平均效应值”分别为 0.696 (99% 的置信区间为 [0.667, 0.723])、0.610 (99% 的置信区间为 [0.575, 0.643]) 和 0.211 (99% 的置信区间为 [0.159, 0.263])。在 2001 年, Djokic 等对 100 个开源的 Java 系统和 100 个开源的 C++ 系统进行元分析,调查 CBO、RFC 和 NOM 之间的相关性^[8]。元分析结果表明,它们之间存在中等到大的相关,由此推测可能存在共线性。特别地, Djokic 等进行了异质性检验,结果表明不存在“异质性”。在 2005 年, Succi 等对 Djokic 等在 2001 年的工作进行扩充,附加调查 NOC 和 DIT 在软件系统中的实际值^[9]。固定效应模型的元分析结果表明,它们的最大值都小于 6。

3.2.4 UML 类图度量与易维护性

在 2008 年, Manso 等对 5 个实验研究进行元分析,调查两个问题^[10]: (1) UML 类图的规模/结构复杂度度量与认知复杂性间的相关性; (2) UML 类图认知复杂性与它们的可理解性/可修改性间的相关性。其中,每个实验研究上的“可观察效应值”为斯皮尔曼相关系数。固定效应模型元分析结果表明: (1) 许多 UML 类图的结构复杂度度量/规模度量都与认知复杂性显著相关,其中关联和泛化关系与认知复杂性具有较强的相关性; (2) 认知复杂性与可理解性/可修改性都具有显著的相关性,其中与可理解性的相关性更强。

3.2.5 结对编程方法的有效性

在 2009 年, Hannay 等对 18 个实验研究进行元分析,调查结对编程的有效性,包括对代码质量、开发时间和开发工作量的影响^[11]。其中,大多数实验研究用通过的测试用例数或代码复杂度来度量代码质量,用所有任务或达到一定质量标准的任务的总完成时间度量开发时间。在每一个研究中,“可观察效应值”为标准化的平均值差。Hannay 等同时采用固定和随机效应模型进行分析,结果表明: (1) 结对编程对代码质量有比较小的正面影响; (2) 对开发时间有中等的正面影响; (3) 对开发工作量有显著的负面影响。更进一步, Hannay 等根据开发人员的类型(学生和专业人员)和比较的方式(个体和团队)将实验研究进行分组,然后进行元分析。在专业人员组中,结对编程对代码质量和开发时间没有显著影响;在按团队比较的组中,结对编程对代码质量、开发时间和开发工作量都没有显著影响。特别地, Hannay 等做了一个更详细的分析,结果发现: (1) 当编程任务的复杂性比较低时,结对编程比传统的方式更快,但得到的代码质量低; (2) 当编程任务的复杂性比较高时,结对编程比传统方式能得到更高质量的代码,但要花费更多的工作量。Q 检验的结果表明其具有较高的“异质性”。

3.2.6 测试驱动开发方式的有效性

在 2012 年, Rafique 等对 27 个相关研究进行元分析,调查测试驱动的开发(TDD)方式对代码质量和生产率的影响^[12]。元分析结果表明, TDD 对代码质量具有小的正面影响,对生产率几乎没有影响。同时,他们分析了开发者经验和任务规模对元分析结果的影响,发现任务规模和质量改进的强度间存在显著的正相关关系。分组后的元分析结果表明,与学术界相比: (1) 工业界中 TDD 对质量有更大的正面影响,

这有可能是由于工业界的开发人员具有更高的开发经验和更大的任务规模造成的; (2) 工业界中 TDD 对生产率具有更大的负面影响,这有可能是更有经验的开发人员通常把更多的工作量花在测试和重构等 TDD 相关活动上造成的。Hannay 等人利用 Q 和 I^2 进行了异质性检验,发现参与元分析的实验研究具有较高的“异质性”。

3.3 问题与展望

在软件工程中,利用元分析方法综合多个实验研究的结果得到可靠的实验结论具有非常重要的意义。尽管元分析方法已经得到人们的重视并在软件工程领域内逐步得到应用,但还有许多问题有待解决。

元分析结论的可靠性和统计效力问题。最近的研究结果表明,为达到 95% 的可靠性(即不犯第一类错误的概率)和 80% 的统计效力(即不犯第二类错误的概率),参与元分析的实验研究数目应该大于等于 8,在这些实验研究中总的实验对象数目应该大于等于 160^[13]。然而,先前的一些元分析研究只使用了 5 个或者更少的实验研究。将来的元分析研究应使用较多的实验研究和实验对象,以确保元分析的可靠性和统计效力。

元分析中实验研究的原始数据的可用性问题。元分析通常使用公开发表的论文中报道的结果进行分析,本质上是采用“二手资料”进行研究,一般得不到原始实验数据。特别地,许多实验研究中报道的指标并不一样。为此,在进行元分析时,需要将它们转换为相同形式的“可观察效应值”。由于得不到原始实验数据,“可观察效应值”有可能受数据预处理方式等许多不明因素的影响,进而可能影响元分析结果的正确性。在软件工程领域,应尽可能使用开源软件中的数据进行元分析,以解决此问题。

高级的元分析方法使用问题。现有的元分析研究一般只使用简单的元分析方法,例如对相关系数进行元分析、对实验组和对照组指标差值进行元分析等。然而,在软件工程领域,许多实验研究使用较高级的数据分析方法,例如主成份分析、线性回归和 logistic 回归等。在这些情况下,非常有必要使用高级的元分析方法对主成份进行元分析或者对回归方程进行元分析,以得到相应的元分析结果。

结束语 软件工程领域内研究人员正日益重视元分析技术,并将其积极应用到各种软件工程主题中,以便将多个实验结果进行合并得到可靠的实验结论。然而,当前只应用了简单的元分析技术,尚需要应用更高级的元分析技术,以更好地解决实证软件工程中的问题。

参 考 文 献

- [1] Briand L, Arisholm E, Counsell S, et al. Empirical studies of object-oriented artifacts, methods, and processes; state of the art and future directions [J]. *Empirical Software Engineering*, 1999, 4(4): 387-404
- [2] Borenstein M, Hedges L V, Higgins J P T, et al. *Introduction to meta-analysis* [M]. John Wiley & Sons, Ltd., 2009
- [3] Mair C, Shepperd M. Making software cost data available for meta-analysis [C] // *Proceedings of the 8th International Conference on Empirical Assessment in Software Engineering*. 2004
- [4] Sjøberg D, Dybå T, Jørgensen M. The future of empirical methods in software engineering research [C] // *FOSE 2007*

- [5] Pickard L M, Kitchenham B A, Jones P. Combining empirical results in software engineering [J]. *Information and Software Technology*, 1998, 40(14): 811-821
- [6] Hayes W. Research synthesis in software engineering; a case for meta-analysis [C]// *Proceedings the 6th International Software Metrics Symposium*. 1999; 143-151
- [7] Succi G, Spasojevic R, Hayes J, et al. Application of statistical meta-analysis to software engineering metrics data [C]// *Proceedings of the World Multiconference on Systemics, Cybernetics and Informatics*. 2000; 709-714
- [8] Djokic S, Succi G, Pedrycz W, et al. Meta-analysis—A method of combining empirical results and its application in object-oriented software systems [C]// *Proceedings of the 7th International Conference on Object-Oriented Information Systems*. 2001
- [9] Succi G, Pedrycz W, Djokic S, et al. An empirical exploration of the distributions of the Chidamber and Kemerer object-oriented metrics suite [J]. *Empirical Software Engineering*, 2005, 10(1): 81-104
- [10] Manso M E, Cruz-Lemus J A, Genero M, et al. Empirical validation of measures for UML class diagrams; A meta-analysis study [C]// *MoDELS Workshops*. 2008; 303-313
- [11] Hannay J, Dybå T, Arisholm E, et al. The effectiveness of pair programming; A meta-analysis [J]. *Information and Software Technology*, 2009, 51(7): 1110-1122
- [12] Rafique Y, Mišić V B. The effects of test-driven development on external quality and productivity; A meta-analysis [J]. *IEEE Transactions on Software Engineering*, May 2012
- [13] Dieste O, Fernandez E, Martinez R G, et al. Comparative analysis of meta-analysis methods; when to use which [C]// *EASE*. 2011; 36-45

(上接第 85 页)

- [2] Li Qiao, Zheng Xiao. Research survey of cloud computing [J]. *Computer Science*, 2011, 38(4): 32-37
- [3] Zhang Yun-yong, Chen Qing-jin, Pan Song-bai, et al. Key security technologies on cloud computing [J]. *Telecommunications Science*, 2010, 26(9): 64-69
- [4] Boccaletti S, Latora V, Moreno Y, et al. Complex networks: structure and dynamics [J]. *Physics Reports*, 2006, 424(4/5): 175-308
- [5] Milo R, Shen-Orr S, Itzkovitz S. Network motifs; simple building blocks of complex networks [J]. *Science*, 2002, 298(5594): 824-827
- [6] Shandilya S G, Timme M. Inferring network topology from complex dynamics [J]. *New Journal of Physics*, 2011, 13(1): 1-11
- [7] Karsai M, Kivela M, Pan R K, et al. Small but slow world; how network topology and burstiness slow down spreading [J]. *Physics Review E*, 2011, 83(2): 1-4
- [8] Larremore D B, Shew W L, Restrepo J G. Predicting criticality and dynamic range in complex networks; effects of topology [J]. *Physical Review Letters*, 2011, 106(5): 1-4
- [9] Perotti J I, Billoni O V, Tamarit F A, et al. Emergent self-organized complex network topology out of stability constraints [J]. *Physics Review Letters*, 2009, 103: 108701-108704
- [10] 吕琳媛. 复杂网络链路预测 [J]. *电子科技大学学报*, 2010, 39(5): 651-661
- [11] 王文强, 张千明. 链路预测的网络演化模型评价方法 [J]. *电子科技大学学报*, 2011, 40(2): 174-179
- [12] Jiang Wen-ke, Tan Yang-hong. Chaotic time series forecasting research based on the combination network [J]. *Computer Engineering & Science*, 2010, 32(8): 134-137
- [13] Zhu Gui-mei, Yang Hui-jie, Yin Chuan-yang, et al. Localizations on complex networks [J]. *Physical Review E*, 2008, 77(6): 1-9
- [14] 汪小帆, 李翔, 陈关荣. 复杂网络理论及其应用 [M]. 北京: 清华大学出版社, 2006
- [15] Zhao Kang, Kumar A, Harrison T P, et al. Analyzing the resilience of complex supply network topologies against random and targeted disruptions [J]. *IEEE Systems Journal*, 2011, 5(1): 28-39
- [16] Lin Chun-nan, Hsieh K L, Roan J, et al. The application of structural holes theory to supply chain network information flow analysis [J]. *Information Technology Journal*, 2011, 10(1): 146-151
- [17] Tsea Y K, Tana K H. Managing product quality risk in a multi-tier global supply chain [J]. *International Journal of Production Research*, 2011, 49(1): 139-158
- [18] Waters J, Ceruti M G. Modeling and simulation of information flow; a study of infodynamic quantities [C]// *Proc of the 15th International Command and Control Research and Technology Symposium (ICCRTS 2010)*. Santa Monica, CA; Space and Naval Warfare Systems Center Pacific San Diego Ca, 2010; 1-35
- [19] Waters J, Ceruti M, Patel R, et al. Information-velocity metric for the flow of information through an organization; application to decision support [C]// *Proc of the 14th International Command and Control Research and Technology Symposium (ICCRTS 2009)*. Washington, DC; Space and Naval Warfare Systems Center Pacific (SSC Pacific), 2009; 1-46
- [20] Masucci A P, Kalampokis A, Eguiluz V M, et al. Extracting directed information flow networks; an application to genetics and semantics [J]. *Physical Review E*, 2011, 83(2): 1-6
- [21] Kimble H J. The quantum Internet [J]. *Nature*, 2008, 453: 1023-1030
- [22] Ceruti M G, Rubin S H. Infodynamics; analogical analysis of states of matter and information [J]. *Information Sciences*, 2007, 177(4): 969-987
- [23] Dodson C T J. Some illustrations of information geometry in biology and physics [J]. *Handbook of Research on Computational Science and Engineering; Theory and Practice*, 2011, 2: 1-27
- [24] Kowalski A M, Martin M T, Zunino L, et al. The quantum-classical transition as an information flow [J]. *Entropy*, 2010, 12(1): 148-160
- [25] Alvim M S, Andrés M E, Palamidessi C. Probabilistic information flow [C]// *Proc of the 25th Annual IEEE Symposium on Logic in Computer Science*. Washington, DC; IEEE Computer Society, 2010; 314-321
- [26] Papana A, Kugiumtzis D. Detection of directionality of information transfer in nonlinear dynamical systems [J]. *Topics On Chaotic Systems*, 2009; 251-264
- [27] Sinatra R, Gómez-Gardeñes J, Lambiotte R, et al. Maximal-entropy random walks in complex networks with limited information [J]. *Physical Review E*, 2011, 83(3): 1-4