

6-2-2020

Agent Modeling of Web Users Orientating Simulation of Internet Public Opinion

Chuan Ai

National University of Defense Technology, Changsha 410073, China;

Bin Chen

National University of Defense Technology, Changsha 410073, China;

Liu Liang

National University of Defense Technology, Changsha 410073, China;

Xiaogang Qiu

National University of Defense Technology, Changsha 410073, China;

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research, Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Agent Modeling of Web Users Orientating Simulation of Internet Public Opinion

Abstract

Abstract: People are individuals with independent behavior in society, and have important influence on the occurrence and development of public opinion emergencies. At the same time, the group composed of individuals is the disaster-bearing body of public opinion events. Therefore, the use of simulation modeling means to explore the occurrence and development of public opinion Mechanism, for the public opinion emergency response measures to provide decision support, the individual behavior of the modeling simulation is particularly important. Based on the analysis of the psychological behavior process of individual information processing, *a complete mechanism model was put forward which could reflect the process of individual psychological behavior and the corresponding reduction model, and studied the theory and method of multi-source data association and integration, based on the actual data, the individual model of the realization and parameter analysis were completed.*

Keywords

public opinion simulation, character modeling, psychological behavior, mechanism model

Recommended Citation

Ai Chuan, Chen Bin, Liu Liang, Qiu Xiaogang. Agent Modeling of Web Users Orientating Simulation of Internet Public Opinion[J]. Journal of System Simulation, 2017, 29(9): 2221-2226.

面向网络舆情仿真的人物智能体建模

艾川, 陈彬, 刘亮, 邱晓刚

(国防科学技术大学信息系统与管理学院, 湖南 长沙 410073)

摘要: 人是社会中具有自主行为的个体, 对网络舆情突发事件的发生、发展具有重要影响, 同时由个体组成的群体是舆情事件的承灾体, 因此使用仿真建模手段探索网络舆情发生、发展机理, 为舆情事件应急处置措施提供决策支撑时, 对个体行为的建模仿真显得尤其重要。面向网络舆情仿真, 分析了个体信息处理的心理行为过程, 提出了能够反映个体心理行为过程的机理模型, 基于实际数据进行了个体模型的实现、参数分析和仿真实验。

关键词: 舆情仿真; 人物建模; 心理行为; 机理模型

中图分类号: TP391.9 文献标识码: A 文章编号: 1004-731X (2017) 09-2221-06

DOI: 10.16182/j.issn1004731x.joss.201709046

Agent Modeling of Web Users Orientating Simulation of Internet Public Opinion

Ai Chuan, Chen Bin, Liu Liang, Qiu Xiaogang

(National University of Defense Technology, Changsha 410073, China)

Abstract: People are individuals with independent behavior in society, and have important influence on the occurrence and development of public opinion emergencies. At the same time, the group composed of individuals is the disaster-bearing body of public opinion events. Therefore, the use of simulation modeling means to explore the occurrence and development of public opinion Mechanism, for the public opinion emergency response measures to provide decision support, the individual behavior of the modeling simulation is particularly important. Based on the analysis of the psychological behavior process of individual information processing, a complete mechanism model was put forward which could reflect the process of individual psychological behavior and the corresponding reduction model, and studied the theory and method of multi-source data association and integration, based on the actual data, the individual model of the realization and parameter analysis were completed.

Keywords: public opinion simulation; character modeling; psychological behavior; mechanism model

引言

随着互联网技术的飞速发展, 人类社会已经进入到了 web2.0 的网络时代。传统社交网络与新兴网络相互交织, 共同影响人们的生活。使用仿真

建模手段探索网络舆情发生、发展机理, 为舆情事件应急处置措施提供决策支撑时, 对个体行为的建模仿真显得尤其重要。

舆论研究始于社会科学领域, 是传播学的主要研究内容之一^[1]。1948年, Lasswell提出了著名的 5 W 模式^[2], 开创了传播学研究的新范式。社会心理学家们则从人类个体的角度对人们如何形成和改变自己的态度或行为, 如何受他人或社会环境影响等问题开展了一系列的调查研究^[3-4]。



收稿日期: 2017-05-10 修回日期: 2017-07-14;
基金项目: 国家自然科学基金(71673292, 61503402),
国家重点研究发展计划(2017YFC803300), 上海软件
与集成电路专项基金(150312);
作者简介: 艾川(1991-), 男, 四川资阳, 博士生, 研
究方向为系统仿真。

<http://www.china-simulation.com>

• 2221 •

对个体的建模从根本上说是对个体的心理行为进行建模, 由于人本身具有多样性、复杂性等特点, 虽然社会学领域有关人类的心理行为方面的成果已经很多, 但没有一个能够完全反应个体在处理信息时的心理行为过程的模型; 很多模型都仅仅存在于理论层面, 没有考虑如何更多地与实际的数据结合, 从而提高可信度。对于个体人物建模亟待新的理论和方法。

近年来, 受益于计算机和互联网技术的发展, 采集网络舆情传播中大量有关个体行为的数据变得容易, 相关的数据挖掘工作取得了很大的成功。比如网上购物网站的自动推荐系统、新闻客户端的新闻自动推荐系统, 这些实际应用的成功说明通过对大量的个体的数据进行挖掘来分析个体的行为规律是切实可行的。

因此, 本文在拥有大量与网络舆情中个体有关的数据的条件下, 引入对大量数据挖掘行为特征的方法和技术, 以此对大量的数据进行分析建模来构建个体 Agent 模型, 可以提高个体人物建模的可行性和可靠性。

1 相关工作

研究者们综合利用社会动力学、复杂性科学、统计物理学等学科的建模方法, 结合社会心理学研究中得出的个体观点如何形成和个体如何改变自身观点的相关结论, 提出了很多舆论演进模型。其中的一些模型对后续的研究产生了深远的影响。

Galam 等人给出了一种有关舆论演化的重要模型, 称为 Galam 选举模型^[5-6]。模型以局部多数原则(Local Majority Rule)为基本规则, 即在一个局部群体中的个体总是以少数服从多数的方式达成一致意见。

Martins 等开展了一系列应用贝叶斯理论对个体决策过程进行建模的工作^[7-10], 建立了 CODA(Continuous Opinion Discrete Attitude) 模型。其基本思路是: 假设个体能够被观察到的态

度与其所持有的真实观点不完全一致, 群体中的每个个体都根据相邻个体的态度来判断是否体现了邻居个体的真实观点, 并据此调整自己的观点倾向, 进而决定自己所表现出的态度。

舆论演化模型很大程度上概括了前人对舆论演化的分析建模方面的成果, 对演化过程进行了一定程度的归纳和抽象, 对于帮助我们了解舆论演化的本质有很重要的意义。同时, 这些工作可以指导我们构建基于数据的个体模型, 在基本的结构上可以借鉴这些成果。然而, 在基于数据的个体建模上这些成果无法满足我们的需求, 问题主要在于如何去挖掘实际数据中的个体属性, 如何将个体的行为归纳成条理清晰方便计算机实现的行为集合, 以及如何使用合理的算法来训练模型这三个方面。这也是本文重点探索的问题。

2 模型构建

对个体的心理行为进行建模需要一个清晰的、逻辑严密的体系结构, 但目前为止, 各种理论的研究工作通常针对特定的案例, 并没有形成通用的可以扩展的理论和方法, 很难形成完备统一的整体, 为模型的建立带来了很大的困难。本节在领域专家指导下构建了个体信息处理完整机理模型, 它基于心理学领域理论和经验, 完整反映了个体在接收、处理和传播在线信息的心理行为。

• 个体机理模型框架

基于网民的媒介素养特征, 将其信息处理的认知和传播过程抽象为两个阶段: 处理与否的判断、自身观点形成。详细的过程如图 1。

网民个体处理信息的前提是能接收到信息, 该模型的第一步就是判断能否接收到相关信息。如果分析网络数据发现该个体以往倾向于关注某些个体, 同时这些个体有发布信息, 该个体就能通过网络平台获取信息, 反之则认为无法接受到信息。收到信息之后, 网民是否处理信息会受到心理距离和心理认知资源的综合作用影响。对于是否传播消息也可以简单设定为如果网民有对信

息进行加工, 有自己的思考, 就会传播信息。反映在模型中就是, 如果模型中判定该个体处理信息, 并且通过观点决策过程得到了自身新的观点, 那么该个体自然就要发表观点, 也就是传播信息。

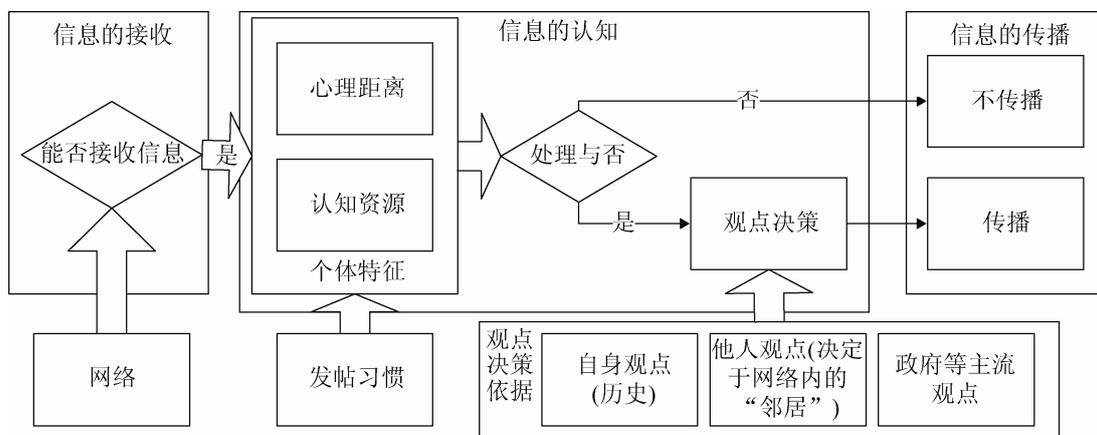


图 1 网民信息处理框架
Fig. 1 Framework of web users' information processing

信息传播阶段主要内容是对网民个体的传播影响力进行刻画, 不同于传播意愿, 其传播能力是自身的影响力大小, 是他的信息能够被多少人所接受。这主要由网络的情况决定。

• 个体状态特征提取

本节介绍心理距离、认知资源计算方法。

根据网民的心理距离的定义, 如果网民对事件不感兴趣或者没有利益关系, 即心理距离较远, 那么网民处理信息的可能性就比较低; 如果对事件很感兴趣, 那么处理事件的愿望就很高。不考虑具体的事件类型或者具体是哪个事件, 可以从另一个层面上去考虑心理距离。社交网络中存在有大量粉丝的大 V, 很多情况下大 V 的言论对个体的影响巨大, 可以从这个角度考虑个体到个体之间的类似于心理距离的一个指标, 即心理亲近度, 这个指标用于反映个体与个体之间的亲密程度。

$$PD_{ij} = N_{t_{ij}} / N_i \tag{1}$$

$$\begin{cases} \delta_{ij}^1 = 1; & PD_{ij} < PD \\ \delta_{ij}^1 = 0; & \text{else} \end{cases} \tag{2}$$

式中: PD_{ij} 反映个体 i 到个体 j 的心理亲近度; N_i 是个体 i 发帖状态或者消息的次数; $N_{t_{ij}}$ 是个体 i 转发 j 的状态或者消息的次数, 值越大说明个体之间越

亲近, 值越小说明个体之间越疏远。

根据网民的认知资源的概念, 事件与自身相关程度比较高的情况下, 个体可能会根据当前的时间和精力的情况来决定是否处理信息、传播信息。定义具体的计算规则如下: 统计该个体每月发消息总量 Num , 计算平均每天发消息数量 $AvrNum$ 。可以设置 $AvrNum$ 即为认知资源容量。

$$AvrNum = Num / 30 \tag{3}$$

$$p = (AvrNum - Num_{now}) / AvrNum \tag{4}$$

Num_{now} 是当前当天已处理消息数量。

3 参数计算

考虑行为决策模型, 其根本是计算出个体接收信息后在多种条件下对应的转发和不转发分别对应的概率。求得这个概率就完成了这个模型的实现工作。

$$p_{mn} = p(C_m, O_n) \tag{5}$$

$$m \in [0, M], n = 1, 2, 3, \dots \tag{6}$$

式中: C_m 表示第 m 种条件; M 为条件总数, $C_m = (time, p_d, p_s)$, 即个体处于不同的时间、心理距离、心理资源, 是条件; O_n 是结果, 包含转发和不转发 2 种。

$$p_{m0} + p_{m1} = 1, m \in [0, M] \tag{7}$$

对于每个个体所有结果的转移概率之和为 1。

- 数据准备

数据预处理过程将数据中不相关字段去除，留下转发记录，字段及含义如表 1 所示：

$$\text{Rcod}_1 = \langle \text{date, time, id, srceid, text} \rangle \quad (8)$$

表 1 预处理结果的五个字段的解释

Tab. 1 Five parts' explanation of pretreatment results

字段	内容	释义
date	日期	当前微博信息发出日期
time	时间	当前微博信息发出时间
id	用户账号	发出该微博的用户
sourceid	用户账号	微博转发来源
text	微博内容	原始微博内容

每条记录为一次转发动作，预处理结果为该动作的五个信息。计算结果中包含了大量用户的转发活动信息。

这里条件转移概率的计算采用了 3 批不同的数据。第 1 批为 2012-04 的微博采样数据；第 2 批为 2012 年上半年的微博采样数据；第 3 批为 2012 年整年的采样数据，如表 2 所示。

表 2 不同批次数据预处理结果的数据量

Tab. 2 Data amount of pretreatment results in different time

批次	持续时间	处理结果/ kb
1	2012 年 3 月	82 339
2	2012 年上半年	461 455
3	2012 年全年	854 480

数据的二次处理是将第一次处理的结果再次分析计算，其目标是找到每一个存在关系的用户之间每次转发和不转发的决策，以及该决策下的条件。举例如下，用户 A 和 B 之间存在好友关系，那么认为用户 A 的消息 B 一定能收到，转发或者不转发是结果，A 发消息的时间，B 到 A 的心理距离、B 此时的心理资源状态作为条件，作为一条“条件-行为结果”记录，写入二次数据处理结果。

$$\text{Record}_2 = \langle C_m, O_n \rangle \quad (9)$$

- 条件转移概率计算

经过数据的预处理和二次预处理，得到了“条件-行为结果”记录。用极大似然估计的方法求解个体的条件转移概率。

对用户 A，假设条件 m 下转发的概率为 p ，相应的不转发的概率为 $1-p$ 。统计转发次数为 l ，不转发次数为 n 。

似然函数为：

$$L(p) = p^l \times (1-p)^n \quad (10)$$

求出使 $L(p)$ 最大的 \hat{p} 即为该用户在此条件下的转移概率。不同的用户个体条件转移概率应当是不同的。计算出的所有结果可以进行分析，进而发现条件与结果的关系。

到此降阶模型的个体条件转移概率计算完成，模型参数的赋值问题也就解决。

4 模型仿真验证实验

本节通过设计试验来验证该模型中行为决策的准确性。用 2012 年上半年的第一批用户的数据作为训练集，对模型参数进行赋值，然后基于该模型预测这些用户 7 月的行为，以 2012 年 7 月同一批微博用户的数据作为对照，验证其行为预测的准确性。

数据处理之后结果如图 2 所示，每个用户都有很多条记录。特别注意，训练集和预测集记录的都是个体收到其他个体的消息后，在当时状态特征下(转发与否)的行为记录。训练集中出现了两个用户之间有转发关系，或者一个用户收到另一个用户的消息，但并没有转发，都说明这 2 个用户之间有消息传递，说明其中一个用户收到了另一个用户的消息。

如图 3 具体的实验步骤为：

(1) 准备 2012 年上半年数据和 7 月数据，进行数据预处理和二次处理。

(2) 以 2012 年上半年的数据，计算出这一批用户的条件转移概率。

(3) 基于 2012 年 7 月同一批用户的数据中出现的用户选择行为(转发或不转发)的条件，预测

其行为转发或不转发)。

(4) 统计每个用户的行为预测准确率, 画出个体转发关系匹配概率、个体行为预测准确率、关系匹配情况下行为准确率分布情况并计算平均准确率。

(5) 得出实验验证结论。

2012年上半年用户A数据二次处理结果

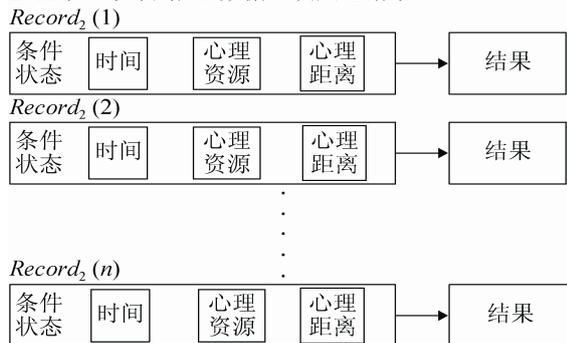


图 2 数据预处理结果

Fig. 2 Data preprocessing results

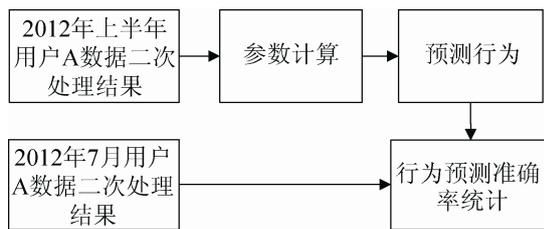


图 3 实验步骤示意图

Fig. 3 Experimental steps

5 实验结果

经过数据准备和条件转移概率计算, 对 7 月中用户的选择行为进行预测, 并统计实际数据中行为选择。对于每个个体统计 3 个数据, 分别是, 个体总体行为数 N_i^t , 预测集中消息传递关系在训练集中出现过的次数 N_i^r , 行为预测准确次数 N_i^p 。

统计完成之后, 对结果再进行处理。

个体转发关系匹配概率: $p_i^r = N_i^r / N_i^t$ 。

个体转发行为预测准确率: $p_i^p = N_i^p / N_i^t$ 。

个体转发关系匹配情况下预测准确率:

$$p_i^{pr} = N_i^p / N_i^r$$

统计出所有个体的 p_i^r , p_i^p , p_i^{pr} 之后, 分别画出其分布图, 如图 4~6 所示。

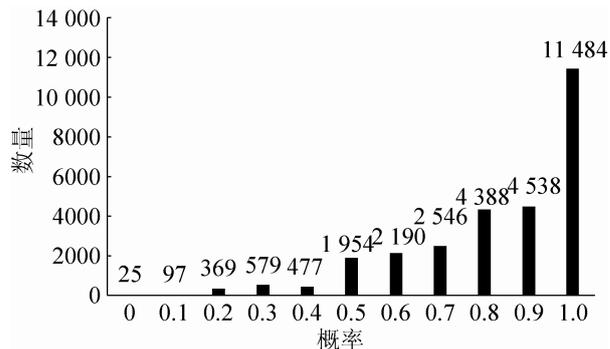


图 4 个体转发关系匹配概率分布

Fig. 4 Individual forwarding relation matching probability distribution

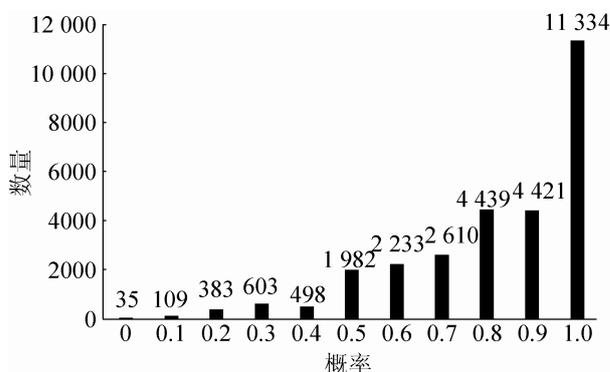


图 5 行为预测准确率分布图

Fig. 5 Behavior prediction accuracy distribution

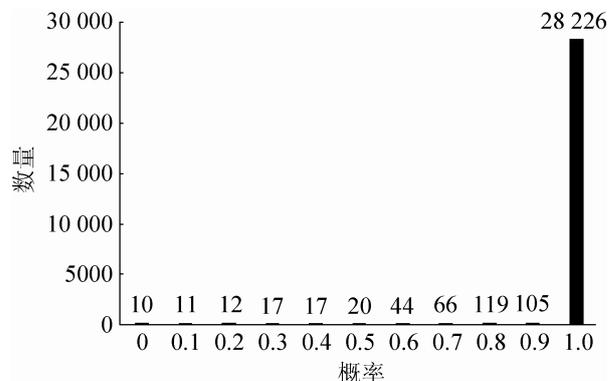


图 6 转发关系匹配时行为预测准确率分布

Fig. 6 Distribution of behavior prediction accuracy when forwarding relation matching

分析图 4 可以看出, 个体转发关系匹配概率的分布情况为概率越大的数量越多, 说明 12 年 7 月出现可能转发的情况基本都在 12 年上半年出现了。从用户全体而言, 匹配率达 86%。

分析图 5 可以看出, 行为预测准确率分布情况也是准确率的数量比较大, 说明行为预测还是

比较准确的。统计所有个体行为预测整体准确率达 84.73%。

行为预测准确率分布与个体转发关系匹配概率分布情况基本一致,这里可以猜测,行为预测准确率受个体转发关系的匹配情况影响很大。图 6 反映的是个体在关系匹配的情况下行为预测准确的概率分布情况。具体解释为,如果预测集中出现了某条件下用户 A 转发或不转发 B 的记录,同时在训练集中也有这样的记录,那么预测的准确性非常高。对于个体总体而言,在训练集中有记录的情况下,预测结果的准确率达 98.4%。

6 结果分析

通过实验,得到了图 4~6,总的来说,仿真的行为预测准确率比较高,能达到 80% 以上,足以支持网络舆情仿真系统的构建。

对于模型的验证结果中较高的准确率需要进行讨论,有几点需要说明。

首先这里得到的准确结果,是确定了当时的条件,包括时间、心理距离和心理资源的情况下得到的,其大前提是我们在数据中已经获知个体在此时已经做出了决定,要么转发,要么不转发。转发和不转发的选择在仿真中一般还会有何时收到消息,何时转发的问题。而实验验证的结果无法说明该模型能做到正确反映个体什么时候看到消息,什么时候转发。

其次,获得的数据中无法得到用户的好友关系,只能根据用户之间是否存在交互活动判定其有关系。现实中有很多好友之间并没有交互,但是用户之间却能够收到消息,这样训练集和预测集中都缺失了一部分用户收到好友消息而没有进行转发的行为记录。

综合以上两点,模型预测的准确率还需要进一步使用更多数据来测试,离应用到网络舆情仿真还有一段距离。

7 结论

本文在舆情传播领域相关专家的指导下构建了个体信息处理模型。通过对微博数据的分析,获取了用户的行为记录,通过对每个用户的转发关系的统计分析,得到了个体信息处理模型的相关参数,并结合实际数据对模型进行了检验。检验结果表明模型达到了预期的目的,但是离实际的应用还有一段距离,需要更加完整的数据支撑才能实现更贴近真实的行为预测。

参考文献:

- [1] Liu Yun, Hu Yu. News and public opinion [M]. Beijing, China: China Radio and Television Publishing House, 2001.
- [2] Lasswell H D. The Structure and Function of Communication in Society [M]. New York, USA: Harper & Row, 1948: 37-51.
- [3] Myers D G. Social psychology [J]. Social Psychology (S0147-829x), 2011, 156(3): 179.
- [4] Zimbardo P G, Leippe M R. The psychology of attitude change and social influence [M]// The Psychology of Attitude Change and Social Influence. USA: McGraw-Hill, 1991: 354.
- [5] Galam S. Minority Opinion Spreading in Random Geometry [J]. Eur. Phys. J. B (S1434-6028), 2002, 25(4): 403-406.
- [6] Galam S. Sociophysics: A Review of Galam Models [J]. International Journal of Modern Physics C (S0129-1831), 2008, 19(3): No. 409440.
- [7] Martins A C R. Continuous Opinions and Discrete Actions in Opinion Dynamics Problems [J]. International Journal of Modern Physics C (S0129-1831), 2008, 19(4): 617-624.
- [8] Martins A C R. Mobility and Social Network Effects on Extremist Opinions [J]. Physical Review E (S2470-0045), 2008, 78(3): No. 036104.
- [9] Martins A C R. Bayesian Updating Rules in Continuous Opinion Dynamics Models [J]. J. Stat. Mech (S1742-5468), 2009 (2): 57-75.
- [10] André C R Martins, Cleber D Kuba. The Importance of Disagreeing: Contrarians And Extremism In The Coda Model [J]. Advances in Complex Systems (S0219-5259), 2010, 13(5): 621-634.