

一个不删词的命名游戏模型

潘秋惠^{*1,2}, 朱 娅¹, 严凌志¹, 胡 骏¹, 贺明峰²

(1. 大连理工大学 创新实验学院, 辽宁 大连 116024;

2. 大连理工大学 数学科学学院, 辽宁 大连 116024)

摘要: 给出了一个不删词的命名游戏模型. 说话者根据自身词库中词的使用频数, 以频数多被选的可能性大作为选词规则; 听者将听到的词与自身词库中的词对照, 若词库中已有该词, 则将该词的使用频数增加, 否则将所听到的词加到自身词库中. 模型研究表明, 每个个体的使用频数在前三位的词中, 至少有一个词是所有个体共有的, 并且排名前三位的其他词也是使用频数较高的, 这为现实中同一个事物一般有大家接受的命名且往往有不同命名的现象提供了一种形成机制. 在此基础上, 讨论了收敛时间问题.

关键词: 命名游戏; 频数; 收敛时间

中图分类号: O242.1

文献标识码: A

doi: 10.7511/dllgxb201403015

0 引言

作为社会动力学的一个分支, 语言发展演化的研究一直被认为是个复杂的问题^[1]. 命名游戏是在没有人为控制和干预情况下, 在每个个体可以创造新的词汇并有自己的词库的基础上, 通过两两之间就某一目标进行命名对话的方式, 最终形成对该目标的命名. Steels 通过著名的听说实验提出了命名游戏模型^[2-3]. Baronchelli 等提出一个后来作为基本模型的简化模型^[4-5].

近年来, 以 Baronchelli 等提出的模型为基础, 针对规则网络^[6]、小世界网络^[7-9]和无标度网络^[10]等拓扑结构, 在加权人际网络^[11]以及二维网络上个体可移动^[12]等情形下, 研究者对形成唯一命名的收敛时间进行了研究. 在语言演化时, 考虑到人们可能倾向于选择那些受欢迎的词来进行对话, Lei 等^[13]提出一个分别建立在小世界网络和无标度网络上的对词语加权的命名游戏模型, 讨论了统一命名的形成问题.

但在现实社会中, 若一个群体就某种说法已达成一致, 外界的人通过一次成功对话就使该群体中的人完全接受新的说法是武断的也是不太稳妥的, 一般认为, 固有的观念不可能轻易被改变.

为此, 本文建立相应的命名游戏模型, 在不删减词库的条件下, 研究每个个体词语使用频数的变化规律. 对话时作为听者, 若听到的词为新词, 则加到词库中, 若为已有的词, 则增加该词的频数以体现其使用强度, 从而依使用频数高低形成一个有序词库; 作为说话者, 若自己词库中无词, 则从外置词库中任选一个词作为选词, 若自己词库中有词, 则以词库中每个词的使用频数为基础, 以频数的归一化权重作为概率进行选词. 最后讨论系统的收敛时间.

1 模型

设有 N 个个体, 任意两个个体都有机会产生联系. 这 N 个个体发现了一个未知事物并试图通过相互交流对其命名. 每个个体都有自己的词库可以储存词汇. 设整个群体共同拥有一个有 M 个词汇的外置词库, 其中的词不因个体的选择而减少. 当个体作为说话者, 自己词库无词时, 就从该外置词库中随机等可能选择一个词, 置于自己的词库中并作为选词告知听者. 初始时, 所有人的词库均为空. 每一时间步, 只有两个个体的对话行为发生.

对话中个体词库的形成及词的使用频数的变

化规则如下:

(1) 任一时间步 t , 记 t 时刻 i 个体的词库中某词 x 的使用频数为 $w_i^x(t)$, 记 $\Omega_i(t)$ 为 i 个体 t 时刻的词库. 随机选择两个个体 i, j 分别作为说话者和听者, 其中 $i, j \in \{1, 2, \dots, N\}$ 且 $i \neq j$.

(2) 对说话者 i , 若其词库为空, 则随机从外置词库中选择一词 x_0 置于词库, 其中 $x_0 \in \{1, 2, \dots, M\}$, 作为选词发言给听者, 且该词的 $w_i^{x_0}(t) = 1, w_j^x(t+1) = w_j^x(t), x \in \Omega_i(t), \Omega_i(t+1) = \Omega_i(t)$. 若其词库非空, i 个体以 $P_i^x(t)$ 的概率选择 x 词, 发言给听者, 其中

$$P_i^x(t) = \frac{w_i^x(t)}{\sum_{x \in \Omega_i(t)} w_i^x(t)}$$

若其发言用词为 x_0 , 则 $w_i^{x_0}(t+1) = w_i^{x_0}(t) + \delta(x - x_0), \delta(x) = \begin{cases} 0; & x \neq 0 \\ 1; & x = 0 \end{cases}, x \in \Omega_i(t), \Omega_i(t+1) = \Omega_i(t) \cup \{x_0\}$.

(3) 对听者 j , 设听到的词为 x_0 , 若词库中有 x_0 , 则 $w_j^{x_0}(t+1) = w_j^{x_0}(t) + \delta(x - x_0), x \in \Omega_i(t), \Omega_i(t+1) = \Omega_i(t)$; 若词库中无 x_0 , 则 $w_j^{x_0}(t) = 1, w_j^x(t+1) = w_j^x(t), x \in \Omega_i(t), \Omega_i(t+1) = \Omega_i(t) \cup \{x_0\}$.

2 结果分析

将每个人词库中的词汇按频数从高到低进行排序可得有序词库. 初始时, 每个说话者从外置词库中随机选择一个词进行交流. 当某一个词出现在所有人有序词库的前三位时, 认为群体达成共识, 此时称系统收敛. 系统在达到收敛时所需的时间称为收敛时间, 记为 T_c . 当从某个时刻开始, 整个系统中的所有词的频数不再发生变化, 此时称系统稳定.

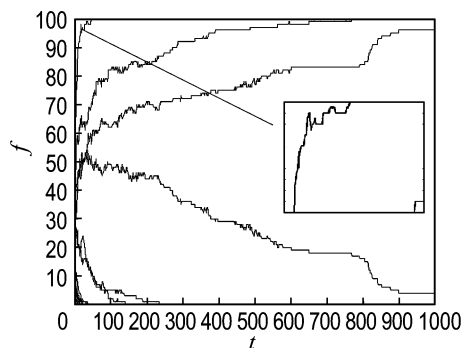
2.1 词的使用频数曲线

取 $N=100, M=100$, 每次模拟 500 万时间步, 每步统计每个人的有序词库前三位的词的使用频数, 选取 3 次的模拟结果如图 1 所示.

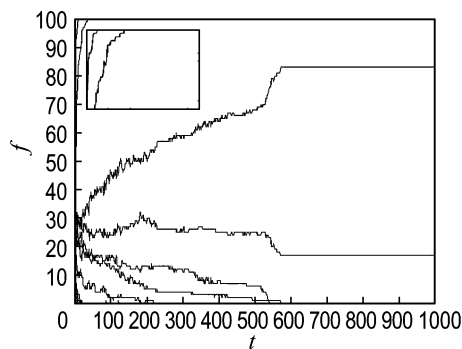
从图 1(a) 可见, 系统收敛较快, 但最终达到稳定的时间较长. 由放大图看出, 排名前三位的词中出现的公共词为 1 个.

从图 1(b) 可见, 有两个词的使用频数迅速上升并在达到 100 后保持稳定, 另一个词的使用频数相对平缓地上升, 使用频数逐步达到稳定, 其余词的使用频数均开始缓慢下降, 其中大部分词最

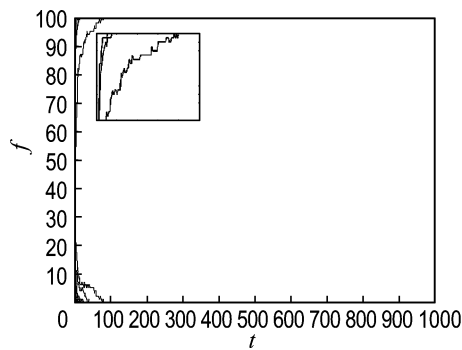
终被淘汰掉. 由放大图看出, 所有人有序词库中排名前三位的词中出现的公共词为 2 个, 即有 2 个词同时出现在所有个体有序词库的前三位中.



(a) 模拟结果 1



(b) 模拟结果 2



(c) 模拟结果 3

图 1 词的使用频数的变化趋势

Fig. 1 Variation tendency of words' use frequencies

从图 1(c) 可见, 系统收敛速度非常快, 前三个优势词在所有人的有序词库中都出现了 100 次, 系统已稳定. 由放大图看出, 排名前三位的词中出现的公共词为 3 个.

综上所述可以看出, 所有人排名前三位的词中出现的公共词至少为 1 个, 有时为 3 个. 使用频数最高的词相对其他词来说有着绝对优势, 且以极快的速度在整个群体中传播开来; 而其他词的变化就较为平缓. 所有曾经出现在每个人的有序词库

前三位的词,其使用频数的变化趋势是同一个方向的,基本没有出现大幅度的先减后增或者先增后减状况.也就是说,一旦变化趋势确定,词汇的流通程度就会始终沿着这个方向前进.

2.2 稳定状态时每个人的有序词库前三位词的使用频数分布

在模拟 500 万时间步停止,图 2 给出了 3 次模拟中每个人有序词库前三位词的使用频数.

从图 2(a)可见,所有人有序词库中排名第一的词包含 4 个不同的词,其中只有 1 个词的使用频数超过了 50,其余的使用频数相对较低;在所有人有序词库中排名前二的词仍为这 4 个,其中在第一幅图中排名第一的词使用频数接近 100,另外 3 个词中的两个与其差距明显减小;在所有人有序词库中排名前三的词包含 5 个不同的词,有一个使用频数为 100,另一个接近 100,一个使用频数极小.所有人有序词库中排名前三的词中只出现了 1 个公共词.

从图 2(b)可见,在所有人有序词库中排名第一的词包含 3 个不同的词,其中只有 1 个词的使用频数超过了 50,其他两个词的使用频数都相对较少,与前者相比差距较大;在所有人有序词库中排名前二的词包含 4 个不同的词,其中在第一幅图中排名第一的词使用频数为 100,另两个词的使用频率也有所增加,同时新增一词,但其使用频数相对前面 3 个词显得极少;在所有人有序词库中排名前三的词包含 6 个不同的词,其中在第一和第二幅图中处于领先地位的 3 个词中的前两个使用频数均达 100,另一个接近 100,其余 3 个词的使用频数相对前面 3 个词显得极少.所有人有序词库中排名前三的词中出现的公共词为 2 个.

从图 2(c)可见,在所有人有序词库中排名前三的词包含 3 个不同的词,其使用频数均达 100,排名前二、第一的仍旧是这 3 个词,且它们的使用频数均相差不大.所有人有序词库中排名前三的词中出现的公共词为 3 个.

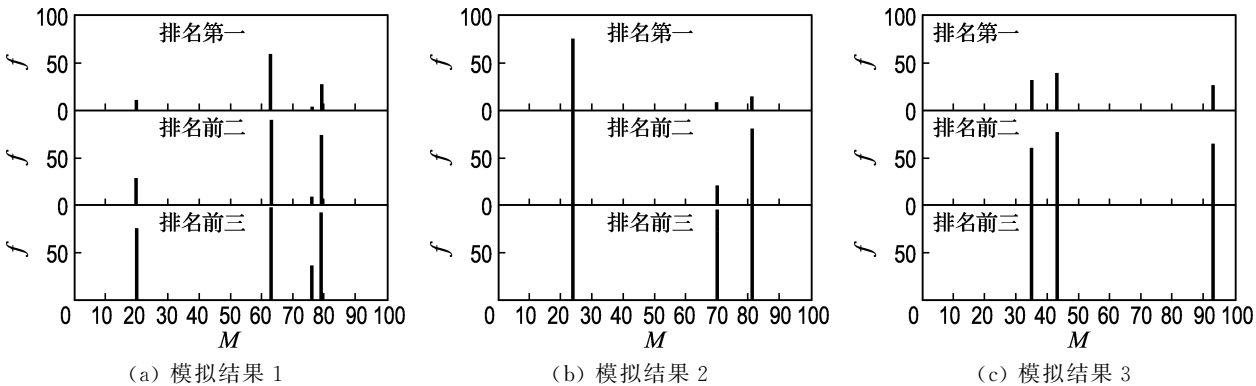


图 2 每个人有序词库前三位词的使用频数分布

Fig. 2 The distribution of the use frequencies of first 3 words in the individual's self-order dictionary

综上所述可以看出,没有一个词能在所有人的有序词库里均排名第一,只有 3 个词能够被整个群体的绝大多数人接受并高频率使用,其余词只能在少数人群中被使用.

2.3 收敛时间

对 $M=100, N=100$, 模拟 1 000 次. 将系统的收敛时间按时间长度进行分组, 分布结果如图 3 所示.

由图 3 可见, 56.8% 的情况下系统在 5 万时间步以内就可收敛, 即所有人有序词库中排名前三的词中出现的公共词至少有 1 个, 19.4% 的情况下系统在 5 万~10 万时间步收敛, 6.5% 的情况下系统在 10 万~15 万时间步收敛, 极个别的情况下系

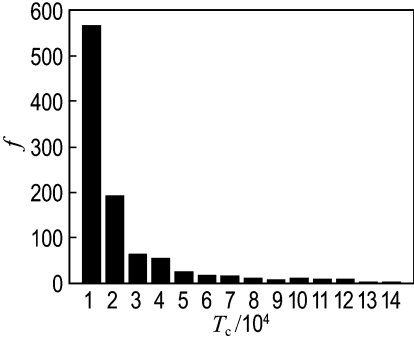


图 3 收敛时间分布

Fig. 3 The distribution of convergence time

统需要 100 万时间步以上才能收敛. 随着收敛时间的递增, 发生的频数迅速下降, 只有个别情况下

系统收敛时间需要 50 万时间步以上. 由此可见, 多数情况下系统在 10 万时间步以内可收敛.

3 结 语

引入加权且不删词的规则后, 研究了能够出现在所有人有序词库中的词, 并重点分析那些在有序词库中排名前三的词的使用频数等. 结果发现, 只要时间足够长, 在所有人有序词库中排名前三的词中出现的公共词至少有 1 个, 有时有 3 个. 也就是说, 在每个人的有序词库前三位中, 就可以有共同的名字. 这体现了名字多样性和统一性. 且有一半以上的情况, 系统可在 5 万时间步内收敛; 80% 以上的情况, 系统可在 15 万时间步内收敛; 收敛时间大于 100 万时间步的情况极少.

参考文献:

- [1] Nowak M A, Krakauer D C. The evolution of language [J]. **Proceedings of the National Academy of Sciences**, 1999, **96**(14):8028-8033.
- [2] Steels L. The synthetic modeling of language origins [J]. **Evolution of Communication**, 1997, **1**(1):1-34.
- [3] Steels L. A self-organizing spatial vocabulary [J]. **Artificial Life**, 1995, **2**(3):319-332.
- [4] Baronchelli A, Felici M, Loreto V, *et al.* Sharp transition towards shared vocabularies in multi-agent systems [J]. **Journal of Statistical Mechanics: Theory and Experiment**, 2006, **2006**:P06014.

- [5] Baronchelli A, Dall'Asta L, Barrat A, *et al.* Topology-induced coarsening in language games [J]. **Physical Review E**, 2006, **73**(1):015102.
- [6] LU Qi-ming, Korniss G, Szymanski B K. Naming games in two-dimensional and small-world-connected random geometric networks [J]. **Physical Review E**, 2008, **77**(1):016111.
- [7] Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks [J]. **Nature**, 1998, **393**(6684):440-442.
- [8] Newman M E J, Watts D J. Scaling and percolation in the small-world network model [J]. **Physical Review E**, 1999, **60**(6):7332.
- [9] Newman M E J. The structure and function of complex networks [J]. **SIAM Review**, 2003, **45**(2):167-256.
- [10] Barabási A L, Albert R. Emergence of scaling in random networks [J]. **Science**, 1999, **286**(5439):509-512.
- [11] Brigatti E. Consequence of reputation in an open-ended naming game [J]. **Physical Review E**, 2008, **78**(4):046108.
- [12] Baronchelli A, Díaz-Guilera A. Consensus in networks of mobile communicating agents [J]. **Physical Review E**, 2012, **85**(1):016113.
- [13] LEI Chuang, JIA Jian-yuan, WU Te, *et al.* Coevolution with weights of names in structured language games [J]. **Physica A: Statistical Mechanics and its Applications**, 2010, **389**(24):5628-5634.

A naming game model without deleting-word process

PAN Qiu-hui^{*1,2}, ZHU Ya¹, YAN Ling-zhi¹, HU Jun¹, HE Ming-feng^{1,2}

(1. School of Innovation Experiment, Dalian University of Technology, Dalian 116024, China;

2. School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China)

Abstract: A new naming game model is presented, which does not include a deleting-word process. In this model, a word is selected out by one speaker with independent possibility according to the use frequencies in the speaker's dictionary. Then, the listener compares the given word with words in his own dictionary. If the word already exists in the dictionary, then the use frequency of the word is increased. If not, the word is added in his dictionary as a new item. The model's experimental results show that there's at least one word in the first 3 high-frequency words of the individual's dictionary, which is shared by all the other dictionaries. Also, an individual's first 3 high-frequency words always own high-frequencies in other dictionaries. The study could provide a formation mechanism to explain why an object would have name acceptable for all persons and probably have different names. Based on this research, the convergence of time in this model is also discussed.

Key words: naming game; frequency; convergence time