



计算机科学

COMPUTER SCIENCE

融合知识图谱的多层次传承影响力计算与泛化研

孔世明, 冯永, 张嘉云

引用本文

孔世明, 冯永, 张嘉云. 融合知识图谱的多层次传承影响力计算与泛化研究[J]. 计算机科学, 2022, 49(9): 221-227.

KONG Shi-ming, FENG Yong, ZHANG Jia-yun. Multi-level Inheritance Influence Calculation and Generalization Based on Knowledge Graph[J]. Computer Science, 2022, 49(9): 221-227.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[时序知识图谱表示学习](#)

Temporal Knowledge Graph Representation Learning

计算机科学, 2022, 49(9): 162-171. <https://doi.org/10.11896/jsjcx.220500204>

[基于 Key-Value 关联记忆网络的知识图谱问答方法](#)

Key-Value Relational Memory Networks for Question Answering over Knowledge Graph

计算机科学, 2022, 49(9): 202-207. <https://doi.org/10.11896/jsjcx.220300277>

[基于自注意力模型的本体对齐方法](#)

Ontology Alignment Method Based on Self-attention

计算机科学, 2022, 49(9): 215-220. <https://doi.org/10.11896/jsjcx.210700190>

[基于知识图谱的层次粒化推荐方法](#)

Hierarchical Granulation Recommendation Method Based on Knowledge Graph

计算机科学, 2022, 49(8): 64-69. <https://doi.org/10.11896/jsjcx.210600111>

[基于自适应注意力机制的知识图谱补全算法](#)

Adaptive Attention-based Knowledge Graph Completion

计算机科学, 2022, 49(7): 204-211. <https://doi.org/10.11896/jsjcx.210400129>

融合知识图谱的多层次传承影响力计算与泛化研究

孔世明¹ 冯永² 张嘉云³

1 重庆文理学院人工智能学院 重庆 402160

2 重庆大学计算机学院 重庆 400044

3 中国检验认证集团重庆有限公司 重庆 401120

(20070114@cqu.edu.cn)

摘要 影响力计算和分析在社交网络、网页重要度评估等领域有着广泛应用。对于有传承链和时间跨度因素的多层次影响力计算,目前尚缺乏较好且通用的解决办法。同时,传播影响力最大化计算是一个 NP 难题,近似算法求解准确度不高且计算复杂。针对上述问题,文中提出了融合知识图谱的多层次传承影响力与泛化算法,实现了传承影响力和传承关系的计算。该算法融合了知识图谱中的广度优先搜索层次计算模型,兼顾时间跨度限制计算传承影响力和传承链;为了优化计算效率,进一步使用深度优先搜索和不同层次加不同权重的策略,只计算前 n 层次的影响力;不仅能很好地计算传承影响力,还可以泛化成各种传播影响力计算模型。在此基础上,文中又提出了通过筛选传播影响力大的节点作为候选节点进行局部最优搜索的传播影响力最大化近似算法,该算法在运行速度和最大传播节点数上都取得了良好的效果。最后,通过多种仿真实验验证了所提方法的有效性。

关键词: 传承影响力计算;传承链计算;知识图谱;传播影响力最大化

中图法分类号 TP391

Multi-level Inheritance Influence Calculation and Generalization Based on Knowledge Graph

KONG Shi-ming¹, FENG Yong² and ZHANG Jia-yun³

1 College of Artificial Intelligence, Chongqing University of Arts and Sciences, Chongqing 402160, China

2 College of Computer Science, Chongqing University, Chongqing 400044, China

3 China Certification & Inspection Group Chongqing Co., Ltd., Chongqing 401120, China

Abstract Influence calculation and analysis are widely used in social networks, web page importance evaluation and other fields. There is still a lack of effective and universal solution for the multi-level influence calculation with inheritance chain and time span factors. At the same time, the calculation of maximizing the propagation influence is an NP hard problem, whose approximate algorithm has low accuracy and complicated computation. In order to solve the above problems, this paper proposes a multi-level inheritance influence and generalization algorithm based on knowledge graph to realize the calculation of inheritance influence and inheritance relationship. The algorithm uses the breadth first search hierarchy computing model of knowledge graph, and takes into account the time span constraints to calculate the inheritance influence and inheritance chain. In order to optimize the computational efficiency, the strategy of depth first search and different levels with different weights is further used to only calculate the influence of the top n levels. The above method can not only calculate the inheritance influence and inheritance chain well, but also can be generalized into various communication influence calculation models. On this basis, this paper proposes a local optimal search similarity algorithm to maximize the propagation influence by selecting the nodes with large propagation influence as spare nodes. It achieves competitive results in running speed and the maximum number of propagation nodes. Finally, the effectiveness of the proposed method is verified by a variety of simulation experiments.

到稿日期:2021-07-13 返修日期:2021-10-19

基金项目:之江实验室开放课题(2021KE0AB01);广西可信软件重点实验室研究课题(kx202006);重庆英才计划创新创业示范团队(CQYC201903167);重庆市技术创新与应用发展专项面上项目(cstc2020jsxc-sbqwX0015);重庆市高技术产业重大产业技术研发项目(2018148208)

This work was supported by the Zhejiang Lab(2021KE0AB01), Guangxi Key Laboratory of Trusted Software(kx202006), Chongqing Talent Plan Innovation and Entrepreneurship Demonstration Team(CQYC201903167), Chongqing Technology Innovation and Application Development Special General Project(cstc2020jsxc-sbqwX0015) and Major Industrial Technology R & D Projects of Chongqing High Tech Industry(2018148208).

通信作者:冯永(fengyong@cqu.edu.cn)

Keywords Calculation of inheritance influence, Inheritance chain calculation, Knowledge graph, Maximization of propagation influence

1 引言

社交网络分析研究具有广泛的现实意义,研究的范畴主要有社交网络传播模型、影响力传播最大化、社交网络社区发现、社交网络的泛化和应用等^[1-5]。社交网络传播模型主要有独立级联、线型阈值模型等^[6-9]。为了便于研究,将各种社交网络模型都抽象成无向图或有向图,将社交网络中的人映射成节点,将人与人之间的关系映射成边^[10]。独立级联模型是由有向图 $G(V, E)$ 及每条边上的影响概率唯一确定的,线型阈值模型就是传播中多个相邻节点共同作用于一个节点,根据其相邻活跃边的权重和是否达到该节点的被激活阈值来确定该节点是否被激活。影响力传播最大化就是找出一个网络中指定数目的种子节点所传播和影响的最大节点数^[11-12]。由于社交网络的传播大都是基于某种概率分布的随机传播网络,求其指定数目的种子节点传播最大化是一个类 NP 问题,或者说是比 NP 问题更难的问题,因此只能采用近似算法求解,普遍采用的是贪心算法和启发式算法。

社交网络可以泛化成现实生活中许多类型的网络,用于分析解决各类实际问题。同样,社交网络中传播的实体多种多样,从具有物理实体的病毒、基因和细菌,到虚拟形式的信息、想法、观念、文化基因以及技术(技艺)等,都可在社交网络中传播^[13]。影响力计算和分析在社交网络、网页重要程度评估等许多方面都有广泛的应用^[14]。对于有传承链和时间跨度因素的多层次影响力计算,目前尚缺乏较好且通用的解决办法。同时,传播影响力最大化计算是一个 NP 难题,近似算法求解准确度不高且计算复杂。针对上述问题,本文提出了融合知识图谱的多层次传承影响力与泛化算法,实现了传承影响力传承关系的计算,主要贡献如下:

(1) 融合知识图谱中的广度优先搜索层次计算模型,兼顾时间跨度限制计算传承影响力和传承链。

(2) 为优化计算效率,进一步使用深度优先搜索和不同层次加不同权重的策略,只计算前 n 层次的影响力。

(3) 提出通过筛选传播影响力大的节点作为候选节点,进行局部最优搜索的传播影响力最大化近似算法,该算法在运行速度和最大传播节点数上都取得了良好的效果。

2 传承影响力计算模型

本文提出了兼顾时间跨度的多层次传承影响力与通用计算模型。该模型的基本思想就是在构建的知识图谱的基础上进行搜索。为了实现计算有顺序的传承链的全谱系计算,采用了广度搜索。在后面的优化算法中,计算前 n 层权重的影响力时,由于不同层次的权重参数会有很大的不同,为了准确地计算传承层次,采用深度搜索。模型中采用了相同算法来计算时间跨度限制的算子和避免图谱节点重复搜索。求最大

传播影响力近似算法是以广度搜索的全谱系计算寻找候选最大传播节点为基础实现的。图 1 给出了本文模型的逻辑结构。

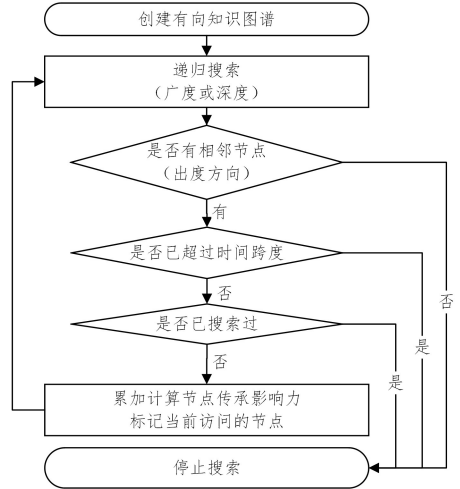


图 1 模型的逻辑结构

Fig. 1 Model logic structure

2.1 全谱系计算传承影响力

在艺术、技术和科学等领域的流派或学派传承时,一个创始人传承若干个第一代传承人,而第一代的每个传承人又传承若干个第二代传承人,第二代传承人又可能传承若干个第三代传承人,这样一直传承下去,传承到第 n 代时,有的传承人可能不会再传承,并且每一代和上一代之间会有一个时间跨度,如图 2 所示。

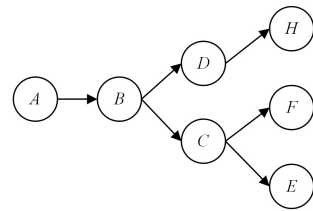


图 2 无环传承

Fig. 2 Loop free inheritance

图 2 中,每个节点代表一个人。除图 2 所示的传承情况外,还可能存在图 3 所示的传承,也就是传承中有环的出现。

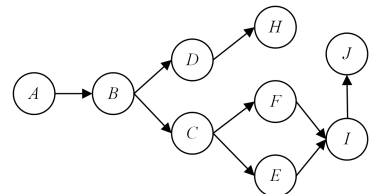


图 3 有环传承

Fig. 3 Ringed inheritance

从图 2 和图 3 来看,传承关系的本质可映射为一个有向图,传承中的每个人就是图中的一个节点,两个节点间的有向边

代表两个人之间的传承关系。于是这个传承社交网络可以表示为有向图 $G=(V,E)$,其中 $V=\{v_1,v_2,v_3,\dots,v_n\}$ 表示图中的点, $E=\{e_1,e_2,e_3,\dots,e_n\}$ 表示图中的有向边。计算某个节点的影响力就是统计从这个节点开始向后传承的节点数,计算式如式(1)所示:

$$ID=\sum_{i=1}^n(\sum_{j=1}^k x),t=M \quad (1)$$

其中, ID 代表影响力, i 代表传承的第几代, j 代表同一代传承中的节点数, x 的值固定为 1, t 代表在设定 M 的时间跨度范围内传承。在 M 时间跨度范围内传承的判定方式如下:

$$t=M(\text{初始状态}) \quad (2)$$

$$t+S \geq N \quad (3)$$

其中, t 初始设定为 M 的跨度时间值,表示所有节点的传承影响力均以该跨度时间标准进行计算, S 代表计算影响力节点的开始传承时间, N 代表后面被传承节点的传承激活时间。只要 N 的值小于等于 t 与 S 的和,就会继续传承下面的节点,直到 N 的值大于 t 与 S 的和为止。如图 3 所示,若在传承中有环存在,则在计算影响力时不重复计算。若从 C 开始统计传承节点,会出现的上分支有 F,I,J ,下分支有 E,I,J ,这样则会重复统计 I 和 J 。为了避免此类情况,已统计过的传承节点需要进行标记,以避免重复统计。

根据上述模型,全谱系计算传承影响力算法的伪代码如下算法 1 所示。

算法 1 全谱系计算传承影响力算法

输入:person(节点),g(传承代数,第 1 次调用只能取 1),y(年限跨度数值加上计算影响力节点的开始传承时间)

输出:影响的节点数,传承链

```

1. Influence_count(person,g,y)
2. a←{}
3. sum=0
4. /* G 代表图对象 */
5. pnode←G(person)
6. if len(pnode) != 0
7. x_sub=0
8. for j=1,2,...,len(pnode) do
9. /* 获取当前子点的 marked 属性的值,初始 marked 属性值为 0,节点被访问后设为 1 */
10. flag1=graph.run(pnode[j].marked)
11. if(flag1 == 0) then
12. graph.run(pnode[j].marked)=1
13. /* pnode[i] 直接影响的所有节点 */
14. pnode1←G(pnode[j])
15. stime=G(pnode[j].time)
16. if stime ≤ y then
17. queue=queue ∪ G(pnode[j])
18. sum=sum+1
19. a=a ∪ pnode[j]
20. end-if
21. end-if

```

```

22. end-for
23. else if(len(pnode) == 0) and (g == 1) then
24. sum←0
25. elseif len(pnode) == 0 and g != 1 then
26. flag2=graph.run(person.marked)
27. if(flag2 == 0):
28. graph.run(person.marked)=1
29. a.append(person)
30. sum=1
31. end-if
32. end-if
33. while queue ≠ ∅ do
34. x,y=influence_count(dequeue(queue),g+1,y)
35. sum=sum+x
36. a=a ∪ y
37. end-while
38. return sum,a

```

算法 1 的时间复杂度为 $O(mn+e)$,其中 n 为节点数, m 为搜索一个节点时处理语句的条数, e 为搜索的边数。该算法不但可以直接计算出每个节点的传承影响力,还可计算出详细的有序传承链。

2.2 层次权重计算传承影响力

在现实世界中,人们综合评定一个艺术家的影响力时,通常具有以下共识:艺术家直接传承的影响力大于间接传承的影响力,间接关系远的传承影响力又小于间接关系近的传承影响力。基于上述共识,为不同传承层次的影响人数分配不同的权重因子,再进行加权来比较艺术家的影响力,更符合实际情况。当人数较多时,直接计算影响力和传承关系,计算复杂度较高,且间接关系越远的传承对提高艺术家影响的贡献逐次衰减,到某一传承代时可以忽略不计。因此,只计算若干代传承影响人数的权重就可逼近真实情况,对降低算法的时间复杂度非常有效。带权值影响力评估值的计算式如式(4)所示:

$$IW=\sum_{j=1}^k(\lambda_j \sum_{i=1}^n x) \quad (4)$$

$$\lambda_j=[\lambda_1,\lambda_2,\lambda_3,\dots,\lambda_n] \quad (5)$$

其中, IW 代表带权重影响力, j 代表传承层次数, i 代表同一传承层次中的实体节点数, λ_j 代表权重系数值, k 的取值一般不大于 5, x 的值固定为 1。层次权重计算传承影响力算法如算法 2 所示。

算法 2 层次权重计算传承影响力算法

输入:person(节点),g(传承代数,第 1 次调用只能取 1),y(年限跨度数值加上计算影响力节点的开始传承时间),max(表示最多可计算到第 max 层)

输出:影响的节点数,传承链

```

1. Influence_count(person,g,y,max)
2. a←{}
3. sum=0
4. /* G 代表图对象 */

```

```

5. pnode←G(person)
6. if len(pnode)≠0 and g≤max then
7. s←{}
8. x_sub=0
9. for j=1,2,⋯,len(pnode)do
10. flag1=graph.run(pnode[j].marked)
11. if(flag1==0)then
12. graph.run(pnode[j].marked)←-1
13. /* pnode[i]直接影响的所有节点 */
14. pnode1=G(pnode[j])
15. /* 当前节点起始受影响时间 */
16. stime=G(pnode[j].time)
17. if G.run(pnode(j))≠∅ and stime≤y
18. s=s∪j
19. sum=sum+1 * λ[g-1]
20. a=a∪pnode[j]
21. end-if
22. end-if
23. end-for
24. for each i∈s do
25. x,y←Influence_count(pnode[1],g+1,y)
26. sum=sum+x * λ[g-1]
27. a=a∪y
28. end-for
29. else if len(pnode)==0 and g=1 then
30. sum=0
31. else if len(pnode)==0 and g<max and g≠1
    then
32. flag2=graph.run(person.marked)
33. if(flag2==0):
34. graph.run(person.marked)←-1
35. a.append(person)
36. sum=1 * λ[g-1]
37. end-if
38. end-if
39. return sum,a

```

算法的时间复杂度为 $O(\mu mn + \beta e)$, u 和 β 的值小于或等于 1, 并且与传承代数的总数有关, 总代数越少, u 和 β 的值就越小。

3 传承影响力计算模型的泛化

传承影响力计算模型除了可用于计算传承影响力外, 也适用于其他社交网络影响力的计算, 如社交舆论传播网络、营销网络、病毒传播网络等, 只需把传承理解为传播即可, 当然这些社交网络大多为随机社交网络。传承网络可被认为是一种固定的社交网络, 可以理解为随机社交网络去掉概率因素退化的社交网络。反之, 传承影响力计算模型也可以泛化为随机社交网络, 就是把节点与节点之间有传承(传播)理解为百分之百的概率变为随机社交网络实际情况下某种概率分布的节点与节点之间的随机概率传承(传播)。即

在计算节点与节点之间传承(传播)时乘上这个随机概率, 根据这个乘积值来确定两个节点间能否传承(传播)。如果实际确定不了真实概率, 则用 $(0, 1)$ 的随机均匀分布概率来代替。

影响力最大化是典型的随机优化问题, 被称为 NP 问题。目前主要依靠寻求近似解的方法来解决影响力最大化问题, 把一个随机社交网络概率去掉后其就退化成一个固定网络, 可以用本文的传承影响力模型来计算 k 个种子节点的最大传播影响力, 用该模型的全谱系算法计算出前 $x * k$ 个传播影响力最大的节点作为候选节点^[15-16]。先选出第一个最大的候选节点作为第一个种子节点, 然后依次在剩下的候选节点中搜索, 找到一个新的节点, 使得相比加入其他节点当前种子节点的集合总传播节点数最大, 这样重复在剩下的候选节点中搜索, 直到搜出指定数量的种子节点数为止。这些种子节点的总影响力就是这些种子节点传播影响力的最大化。这可作为计算传播影响力最大化的一种近似方法, 也可作为其他近似计算传播影响力最大化算法的计算结果的评估对比指标。传承影响力泛化计算最大影响力算法如算法 3 所示。

算法 3 传承影响力泛化计算最大影响力算法(IGMI)

输入: G (图), k (种子节点数), x (倍数)

输出: 种子集合, 种子集合影响的节点和节点数

```

1. /* 求备选前 x * k 个影响力最大的元素 */
2. list_n=∅ /* 存放备选的 x * k 个元素 */
3. for each n∈nodes do
4. sum,list1=influence_count(n,1,y)
5. temp=n,sum,list1
6. if len(list_n)<k * x then
7. list_n=list_n∪temp
8. sorted(list_n)
9. else
10. 在 list_n 中添加比 list_n 中的原有元素影响力更大的元素, 并删去 list_n 中最小的元素
11. end-if
12. end-for
13. start_max 存放种子节点的集合, 首先把影响力最大的第一个节点存进去
14. 取出影响最大的第一个节点所影响的节点并放到 list_max 中
15. for each i∈k-1 do
16. temp_max1=list_max
17. temp_max2=list_max
18. node_x=-1
19. for each j∈list_n do
20. m=len(list_n)-1-j
21. temp_max2=temp_max2 ∪list_n[m][2]
22. if len(temp_max2)>len(temp_max1)then
23. temp_max1=temp_max2
24. node_x=m
25. end-if

```

```

26. end-for
27. list_max=temp_maxl
28. if node_x== -1
29 node_x=len(list_n)-1
30. end-if
31. start_max=start_maxlist_n∪node_x][0]
32. list_n.remove(list_n[node_x])
33. end-for

```

求候选前 $x * k$ 个影响力最大的元素的时间复杂度为 $O(n(mn+e))$, 然后选取 k 个种子节点, 每个种子节点在 $x * k$ 个候选节点中需要搜索 $x * k$ 次, 因此整个算法的时间复杂度为 $O(n(mn+e)+2xk)$, 其中 x 为种子节点数的倍数, x 的值越大, 候选节点数就越多(但 $x * k$ 不能大于总节点数), 计算出来的影响传播节点数就越多, 但 x 的值大到某个数时就不会再增加, 并且算法运行时间也会随候选节点的增多而增加。

4 实验

实验采用了社交网络分析中 neo4j, networkx 和 igraph 这 3 个最流行的知识图谱软件工具和 python3.7, 以及两个数据集(音乐家艺术传承影响数据集 influence_data(2021 美国数学建模大赛 D 题中的数据集)和维基选票网站数据集 Wiki-Vote)。所有实验在 Intel i7-7700HQ(2.80 GHz)、内存 8GB, windows10 操作系统上进行。

4.1 音乐艺术家传承影响力计算

本实验中的传承影响力算法模型在 neo4j 图数据库、python3.7 和 py2neo 包环境下编程实现。采用的验证数据集是音乐家艺术传承影响数据集, 即 influence_data。数据集包括影响者的名字和艺术接受时间, 以及被影响者的名字和艺术被影响起始时间等字段。图 4 给出了 1320 个节点和 1750 个关系用 neo4j 作出的知识图谱。



图 4 1320 个节点和 1750 个关系的知识图谱

Fig. 4 Knowledge graph of 1320 nodes and 1750 relationships

图 5 给出了分别在 2445(6492 条边)、3183(11488 条边)、3526(14988 条边)、5606(42709 条边)这 4 种节点情况下传承 5 层次权重计算影响力和直接计算影响力两种算法的时间对比结果。由于节点数之间相差过大, 不便作图显示, 因此需进行归一化处理, 计算式为 $\lg_{10}(x)/\lg_{10}(\max)$, 然后

乘以 100。权重参数采用 [0.4, 0.25, 0.20, 0.1, 0.05] 这样一组经验参数来计算。

由图 5 可知, 只有在节点数特别大时, 5 层次传承影响力权重计算运行耗费的时间才会明显比全谱系计算传承影响耗费的时间短。因为节点数较少时传承的层次大多都没有超过 5 个层, 所以这时层次权重计算与全谱系计算就很接近。

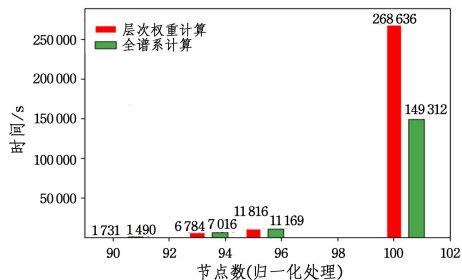


图 5 五层次权重计算和全谱系计算时间对比

Fig. 5 Time comparison of five levels weight calculation and full spectrum calculation

由图 6 可知, 在 2445(6492 条边)、3183(11488 条边)、3526(14988 条边)这 3 种节点上的 3 层次传承影响力权重计算运行耗费的时间, 即使在节点数较少的情况下也比直接全谱系传承影响耗费的时间短。这是因为当节点较多时传承的层次数在大多数情况下都会多于 3。

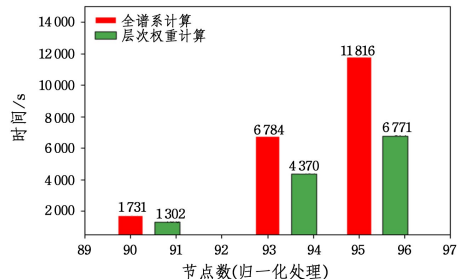


图 6 三层次权重计算和全谱系计算时间对比

Fig. 6 Time comparison of three levels weight calculation and full spectrum calculation

4.2 最大影响力算法对比实验

本文对比实验采用的实验环境是 python3.7, networkx2.5 和 python-igraph0.9.6, 数据集采用的是维基选票网站数据集 Wiki-Vote(有 2184 个节点, 14339 条边), 数据集只有投票者编号和被选举者编号两个字段, 因此算法中忽略了时间跨度因素。

IGMI(多层次传承模型泛化计算最大传播影响力)候选节点的大小为种子节点的 20 倍。用 networkx 建知识图谱实现的算法, 除了 IGMI, 还有 LT-GREEDY(线型阈值模型贪心法)、LT-IMPROVE(线型阈值模型改进)、IC-GREEDY(独立级联模型贪心法), 只有 CELF(成本效益延迟转发)采用 igraph。LT-GREEDY 和 LT-IMPROVE 算法的激活阈值都采用的是 0.5, IC-GREEDY 算法的所有有向边都采用的是 (0,1)之间的随机均匀分布概率除以 2, CELF 算法激活概率

值采用的是 0.08, 核心算法迭代次数为 1000。从实现的底层来看, igraph 构成的图谱运行速度快于 networkx, 因为 igraph 是用 c 编写的, 而 networkx 是用 python 编写的。通过图 7、图 8 的对比实验可以看出, 通过本文提出的 IGMI 计算出 n 个种子节点的最大传播影响力, 分别与 LT-GREEDY, LT-IMPROVE, IC-GREEDY^[17-18] 和 CELF^[19] 进行比较, 相同种子节点下运行时间耗费比另外几种算法都短, 并且比有的算法短很多; 种子节点传播的节点总数同样也比另外几种算法多。其原因在于, 贪心法每一步都要统计还未激活节点的边际效益, 因此时间复杂度很大^[20], 但该算法中的传播节点数相对较多; 而贪心法的改进算法及启发式算法的时间复杂度降低了许多, 但算法中的传播节点数不多; 本文的 IGMI 算法的时间复杂度主要取决于对每个节点的传播数计算和排序, 种子集合的筛选相对于前者可忽略不计, 两者合起来的时间复杂度应远小于或小于贪心法及其改进算法; 而种子集合的挑选本身又是在传播数排名靠前的候选节点基础上筛选出来的。因此, 本文的多层次传承影响力模型泛化的算法是一种比较优秀的计算传播影响力最大化的近似算法。

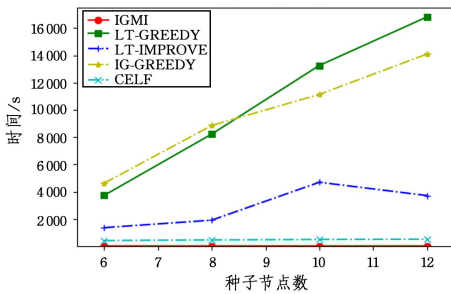


图 7 最大传播影响力各算法运行时间对比

Fig. 7 Running time comparison of algorithms with maximum propagation influence

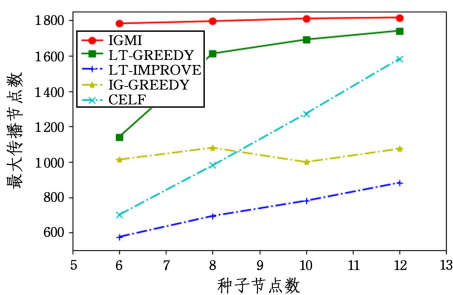


图 8 最大传播影响力各算法传播节点数对比

Fig. 8 Comparison of number of propagation nodes of algorithms with maximum propagation influence

结束语 本文对多层次传承网络的影响力和传承链计算进行了较为深入的分析, 在融合知识图谱和考虑时间跨度等因素的基础上用广度优先搜索算法较好地解决了传承网络影响力和传承链的计算。同时又对这一算法进行了优化, 考虑了不同传播层次的权重因素, 用深度搜索算法实现了更符合实际的权重影响力计算, 经过实验验证, 权重影响力计算的

层次计算到第三层是一个合理层次, 这样既能提高算法的运行速度, 又能较好地评估传承影响力。权重影响力计算到第五层时, 在数据量不是特别大时运行速度和直接计算影响力的运行速度非常接近。

传承网络是一种固定网络, 但与随机概率网络没有本质区别, 它们之间可以通过泛化和退化进行相互转化, 从而把难处理的随机概率网络退化为一种固定网络, 在固定网络上创建模型和设计算法, 可以求其近似解。如本文取前 x 乘以 k 个最大传承影响力的节点作为候选节点来计算 k 个种子节点的最大传播影响力, 算法简单效果却较好。需要改进和提高之处是, 可以引入分布式运算和动态规划算法来提高运算速度; 利用权重计算传播影响力时, 需要优化权重系数取值; 还有就是固定网络泛化为随机概率网络, 并对这种概率值的取值进行探究。

参考文献

- [1] BIAN R, KOH Y, DOBBIE G, et al. Identifying top-k nodes in social networks: a survey[J/OL]. ACM Computing Surveys, 2019, 52(1). <https://dl.acm.org/doi/epdf/10.1145/3301286>.
- [2] HUANG C, JIANG W, WU J, et al. Personalized review recommendation based on users' aspect sentiment[J]. ACM Transactions on Internet Technology, 2020, 20(4): 1-20.
- [3] LIU Q, XIANG B, YUAN N J, et al. An Influence Propagation View of PageRank[J]. ACM Transactions on Knowledge Discovery from Data, 2017, 11(3): 1-30.
- [4] CENTOLA D. The spread of behavior in an online social network experiment[J]. Science, 2010, 329(5996): 1194-1197.
- [5] NEKOVEE M, MORENO Y, BIANCONI G, et al. Theory of rumour spreading in complex social networks[J]. Phys. A, 2007, 374(1): 457-470.
- [6] KEMPE D, KLEINBERG J M, TARDOS É. Maximizing the spread of influence through a social network[C]// Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Washington, DC, USA, 2003: 137-146.
- [7] TANG Y, XIAO X, SHI Y. Influence maximization: near-optimal time complexity meets practical efficiency[C]// Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data. ACM, 2014: 75-86.
- [8] WANG Z X, SUN C C, XI J K, et al. Influence maximization in social graphs based on community structure and node coverage gain[J]. Future Generation Computer Systems, 2021, 118: 327-338.
- [9] NGUYEN H T, THAI M T, DINH T N. Stop-and-stare: optimal sampling algorithms for viral marketing in billion-scale networks[C]// Proceedings of the 2016 International Conference on Management of Data. ACM, 2016: 695-710.
- [10] QU Q, LIU S, ZHU F, et al. Efficient Online Summarization of Large-Scale Dynamic Networks[J/OL]. IEEE Transactions on

- Knowledge & Data Engineering, 2016. <https://ieeexplore.ieee.org/document/7549016>.
- [11] PEDRO D, MATT R. Mining the network value of customers [C]//Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2001:57-66.
- [12] QIU L Q, TIAN X B, SAI S Q, et al. LGIM: A Global Selection Algorithm Based on Local Influence for Influence Maximization in Social Networks[J]. IEEE Access, 2020, 8: 4318-4328.
- [13] CHEN W. Overview and classification of network propagation models[M]//Network Diffusion Models and Algorithms for Big Data. People's Posts and Telecommunications Press, 2020.
- [14] DING Z Y, JIA Y, ZHOU B, et al. Survey of Influence Analysis for Social Networks [J]. Computer Science, 2014, 41(1): 48-53.
- [15] KIANIAN S, ROSTAMNIA M. An efficient path-based approach for influence maximization in social networks[J/OL]. Expert Systems with Applications, 2020, 167(6). <https://www.sciencedirect.com/science/article/abs/pii/S0957417420309064>.
- [16] LI W, ZHONG K, WANG J, et al. A Dynamic Algorithm based on Cohesive Entropy for Influence Maximization in Social Networks [J/OL]. Expert Systems with Applications, 2020, 169(2). <https://www.sciencedirect.com/science/article/abs/pii/S0957417420309350>.
- [17] Social network influence maximization algorithm (linear threshold algorithm and improved algorithm) [EB/OL]. [2018-05-26]. https://github.com/Asia-Lee/Linear_Threshold.
- [18] Influence-maximization (LT and IC) [EB/OL]. [2015-05-19]. <https://github.com/nd7141/influence-maximization>.
- [19] Influence-maximization(CELF)[EB/OL]. [2018-09-07]. https://hautahi.com/im_greedyself.
- [20] QIU L, TIAN X, ZHANG J, et al. LIDDE: A differential evolution algorithm based on local-influence-descending search strategy for influence maximization in social networks[J/OL]. Journal of Network and Computer Applications, 2021. <http://www.sciencedirect.com/science/article/pii/S1084804520304240>.



KONG Shi-ming, born in 1972, master, lecturer, is a member of China Computer Federation. His main research interests include social network analysis, knowledge graph and machine learning.



FENG Yong, born in 1977, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation. His main research interests include big data analysis and data mining, artificial intelligence and big data processing, deep learning.

(责任编辑:喻藜)