Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 10, No. 2, 2024

Study of the Agricultural Applications of the PageRank Algorithm

Tianqi Liu, Mingqiang Gao*

University of Science and Technology Liaoning, Anshan, China
* Corresponding author

Abstract: As a traditional agricultural country, agriculture is the foundation of China's economic and social development. Therefore, promoting the development of agriculture is a necessary way to realize common prosperity. Applying the PageRank algorithm to agriculture can not only liberate the labor force, but also improve the crop yield to achieve economic optimization and promote social development and progress. By exploring the advantages and disadvantages of the PageRank algorithm, based on the concept of graph theory, different weights are assigned to the links according to the attribute characteristics of the links themselves. Then according to the data validation, experimental analysis, and the results of weight size, match and screen the corresponding web pages, and according to the matching degree and importance of the sequential order, and then collect data. Then the relevant knowledge of graph theory is used to illustrate the crop growth environment and growth factors, and the growth status of the target crops is observed in real-time to optimize the growth conditions at any time. Thus, the optimal growing environment is determined, and the yield of crops is improved to promote economic development.

Keywords: PageRank Algorithm, Nodal, Graph Theory, Crops.

1. Introduction

With the rapid development of technology, the Internet of Things (IoT) has become one of the key technologies for global connectivity.[1-4] IoT is a network that connects various physical devices and objects through the Internet for data exchange and intelligent control.[5] Among them, the Internet is a system of interconnected computer networks capable of connecting billions of people around the world, while objects can be biological or non-biological things.[6] In this context, people can connect various devices to the Internet through smartphones, sensors, wireless communications and other technologies to realize functions such as remote monitoring, data collection and intelligent control,[7] which provide people with a more convenient, efficient, and intelligent way of living and working, as well as promote economic growth and social progress.

The development of IoT is based on the need for information sharing and resource utilization. However, based on consumer demand, data needs to be analyzed and optimized to extract valuable information.PageRank algorithm, as a link analysis algorithm, can be used to analyze the connectivity relationship and importance between devices in the IoT so as to optimize the efficiency of data transmission and device control.[8-12] In addition, PageRank algorithm can also evaluate, filter and identify influential nodes in IoT, which can help in rational allocation and management of resources. Applying PageRank algorithm to IoT can achieve intelligent management and optimization of IoT system, and improve the performance and efficiency of the system.

In recent years, with the increasing environmental problems, it has led to many challenges and dilemmas in the field of agriculture.[13,14] For example, climate change has led to an increase in extreme weather events, such as droughts, floods, and storms, which have had a serious impact on crop production and agroecosystems[15]; long-term over-farming, the use of chemical fertilizers and pesticides, as well as industrial emissions and domestic wastes in the process of urbanization have led to soil degradation and contamination, which have seriously affected crop growth and the sustainable use of land.[16] In addition, with the progress of science and technology, it is still facing problems such as rising costs of agricultural production, lower production efficiency, and increased pressure of market competition.[17] Therefore, it has become an important challenge to circumvent extreme environmental impacts, improve agricultural productivity and achieve sustainable development.

The aim of this work is to improve productivity and save labor cost by applying PageRank algorithm to agricultural production. The theoretical routing algorithm that describes the shortest path in graph theory for crop growth is used to determine the optimal conditions suitable for the growth of the target crop through continuous screening and adjustment. This research plays an important role in efficient crop growth, improving crop yield and promoting social and economic development.

2. PageRank Algorithm

The PageRank algorithm was developed by Google founders Larry Page and Sergey Brin in 1997 when they were building an early prototype of the search system. The link analysis algorithm is used to rank websites in Google searches and is a measure of how important a web page is. The PageRank algorithm was used in commercial acquisitions and was so successful that it has attracted a lot of attention from the academic community.

The PageRank algorithm calculates the PageRank value of a web page by the number of inbound and outbound links to the page, with the underlying assumption that more important sites are likely to get more links from other sites. Inbound links are links to the current web page and outbound links are links from the web page. The PR of a web page is the sum of the inbound link weights of all links to the web page, and the PR of the current web page is given equally to each link

chained out, then there:

$$PR(u) = \sum_{v \in I_u} \frac{PR(v)}{N_v} \quad (1)$$

The two assumptions on which the PageRank algorithm is based are: (1) A web page is an important web page if it is referenced by many web pages, known as the quantity assumption. (2) a web page is also an important web page if there are high-quality web pages (authoritative web pages) pointing to it, known as the quality assumption, and the core formula[18] of the PageRank algorithm is as follows:

$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(P_j)} \quad (2)$$

Where the $PR(p_j)$ is the PageRank value of web page p_i, $M(p_i)$ is the set of web pages chained into p_i, $L(P_j)$ is the number of web pages chained out of web page P_j, N is the number of all the web pages in the set, and d is the damping factor, which is usually taken as 0.85[19].

To summarize their advantages and disadvantages are:

Advantage: PageRank is a kind of static arithmetic, independent of the query, through offline calculation to obtain the PR value of all web pages; it can effectively reduce the online query when the amount of calculation is much less, greatly reducing the query response time.

Disadvantages: PageRank analyzes the importance of a web page based only on the degree of links to the page, ignoring factors such as topic relevance and context, which results in a not-so-reasonable ranking result. For example, a new web page, even if the upstream link of the web page will not be much attention, the order will be higher and lower, unless it is a certain sub-site. However, the application of this algorithm to smart agriculture will also be a major breakthrough in agricultural development.

3. The Development of IoT Technology and The Current State of Agriculture Today

With the rapid development of agricultural information technology, China has put forward the concept of smart agriculture, which gradually applies the IoT technology to the field of agricultural production, carries out real-time monitoring of the planting environment, achieves the scientific, automated and intelligent fine management of agricultural production, reduces the management costs of personnel, and improves the yield and quality of planted crops.[20] Among them, NB-IOT and Lora and other communication methods are more suitable for the transmission of relevant monitoring data, with the development of Huawei Cloud, Ali Cloud and other related servers, as well as the development of MQTT message queuing telemetry transmission protocols and NodeMCU open source and fast hardware platforms, so that the cost of communication facilities related to smart agriculture is getting lower and lower, which is more suitable for modern agricultural production.

Nowadays, as far as the western region is concerned, the geographical conditions are harsh, the western region has a complex geography and harsh climatic conditions, and most of the area is arid or semi-arid, and agriculture is facing problems such as water shortage and land degradation. Combination of agriculture and animal husbandry: Agriculture in the western region is based on the combination of agriculture and animal husbandry, with the planting of food crops and pasture as the mainstay, and the development of animal husbandry at the same time. Compared with the level of developed areas in China is too low, so the development of intelligent agriculture is far-reaching.

4. PageRank Algorithm and Smart Agriculture

Firstly, the crops growth as a directed graph, simple assumptions inside the relevant factors are light time, temperature, soil humidity, air humidity, soil PH value and other related factors, light time directly affects the temperature of the environment where the crop is located, soil humidity and air humidity, soil humidity affects the soil PH value and so on.

To calculate the rank of each node, the weight of each incoming link must be calculated using the rank of the other nodes. But in order to do that, we have to compute the rank of every other node, then the rank of the nodes linked to other nodes, and so on. So, linear algebra is a convenient and efficient way to create a matrix representation of this graph and use iterative vector multiplication to achieve our results.

The Python code for calculating and computing iterations are described as :

```
import numpy as np
def calculate_pagerank(transition_matrix):
    n = len(transition_matrix)
    err_bound = 0.005 # Personally set error bounds, not sure what values Page and Brin used in their implementations.
    v1 = np.ones(n) / n # Initialize v1 as a uniformly distributed probability vector.
    v2 = np.matmul(transition_matrix, v1)
    count = 1

    while count <= 100: # Set the maximum number of iterations to 100.
    if np.linalg.norm(v2 - v1, ord=np.inf) < err_bound:
    # Use the infinite parameter to determine if the error between vectors is less than the error bounds.
    break
        v1 = v2
        v2 = np.matmul(transition_matrix, v1)
        count += 1
    return {'vector': v2.tolist(), 'iterations': count}
transition_matrix = np.array([[0.1, 0.2, 0.7],
                              [0.4, 0.5, 0.1],
                              [0.3, 0.6, 0.1]])

pagerank_result = calculate_pagerank(transition_matrix)
print("PageRank vector:", pagerank_result['vector'])
```

```
print("Iterations:", pagerank_result['iterations'])
```

The powers of the transfer matrix are computed iteratively to approximate the PageRank vector. The iterative computation process will stop when the error between two consecutive iterations is less than a given error bound. Finally, the calculated PageRank vector and the number of iterations are returned.

The main steps including that (1) initialize the variables: first obtain the size of the transfer matrix, then set the error bounds and the initial vector v1 as a uniformly distributed probability vector. (2) Iterative computation: the PageRank vector is computed iteratively using matrix multiplication until the error between two iterations is less than a given error bound or the maximum number of iterations is reached (here set to 100). (3) Return results: return the PageRank vector and the number of iterations.

The within_err_bound function. this function is used to check if the difference between two vectors is within the error bounds. If the absolute value of the difference between the two vectors for each element is less than the error bound, then return True, otherwise return False.

5. Conclusion

PageRank algorithm can be used for the importance assessment of nodes and the identification of key nodes in the management of the agricultural industry chain. By constructing an agricultural industry chain network, taking each link of the industry chain as a node in the network, and then using the PageRank algorithm to sort these nodes, the key links and enterprises in the industry chain can be screened out to provide reference and basis for the optimization and management of the agricultural industry chain. Based on this, the application of the PageRank algorithm in agriculture can be as follows:

(1) Agricultural products market analysis: It can be used to rank enterprises or products in agricultural products market analysis. By constructing the agricultural market network, the competitive relationship between different enterprises or products is constructed into a network, and the PageRank algorithm is used to rank these nodes, so that the leading enterprises or best-selling products in the market can be found, which provides support for the analysis of competition in the agricultural market.

(2) Rural tourism development: it can be used to assess the importance of attractions and predict tourist flow in rural tourism development. By constructing a rural tourism network with links between attractions and utilizing the PageRank algorithm to rank these nodes, the importance and attractiveness of tourist attractions can be found, providing a reference basis for the development and management of rural tourism resources.

(3) Rural e-commerce development: it can be used to rank online stores and recommend products in rural e-commerce development. By constructing a rural e-commerce network with associations between online stores and products, and utilizing the PageRank algorithm to rank these nodes, popular online stores and hot-selling products can be found, providing support for the operation of rural e-commerce platforms and product recommendation.

References

[1] Mouha R A. Internet of things (IoT) [J]. Journal of Data Analysis and Information Processing, 2021, 9(2): 77-101.

[2] Sarker I H, Khan A I, Abushark Y B, et al. Internet of things (IoT) security intelligence: a comprehensive overview, machine learning solutions and research directions [J]. Mobile Networks and Applications, 2023, 28(1): 296-312.

[3] Kaur B, Dadkhah S, Shoeleh F, et al. Internet of things (IoT) security dataset evolution: Challenges and future directions [J]. Internet of Things, 2023: 100780.

[4] Ystgaard K F, Atzori L, Palma D, et al. Review of the theory, principles, and design requirements of human-centric Internet of Things (IoT) [J]. Journal of Ambient Intelligence and Humanized Computing, 2023, 14(3): 2827-2859.

[5] Chen Xiaohong, He Caicai, Chen Yan, Xie Zhiyuan. Internet of Things (IoT)-blockchain-enabled pharmaceutical supply chain resilience in the post-pandemic era [J]. Frontiers of Engineering Management, 2023, 10(1): 82-95.

[6] Madakam S, Ramaswamy R, Tripathi S. Internet of Things (IoT): A literature review [J]. Journal of Computer and Communications, 2015, 3(5): 164-173.

[7] Dou F, Ye J, Yuan G, et al. Towards artificial general intelligence (AGI) in the internet of things (IoT): Opportunities and challenges [J]. 2023, 2309: 07438.

[8] Chen Zijun, Ma Delong, Wang Yishu, Yuan Ye. GPPR:Cross-geo-distributed Personalized PageRank Algorithm [J]. Journal of Software, 2024,35(03):1090-1106.

[9] Berkhin P. A survey on PageRank computing [J]. Internet mathematics. 2005, 2(1): 73-120.

[10] Xing W, Ghorbani A. Weighted pagerank algorithm. In Proceedings. Second Annual Conference on Communication Networks and Services Research, 2004, (pp. 305-314). IEEE.

[11] Boldi P, Santini M, Vigna S. PageRank as a Function of the Damping Factor. In Proceedings of the 14th International Conference on World Wide Web. 2005, (pp. 557-566).

[12] Rogers, I. The Google Pagerank algorithm and how it works [J]. 2002.

[13] Sumberg J, Giller K E. What is 'conventional'agriculture? [J]. Global Food Security, 2022, 32: 100617.

[14] Jiang Meng, Song Yue, Kanwar M K, et al. Phytonanotechnology applications in modern agriculture [J]. Journal of Nanobiotechnology, 2021, 19: 1-20.

[15] Khan N, Ray R L, Sargani G R, et al. Current Progress and Future Prospects of Agriculture Technology: Gateway to Sustainable Agriculture [J]. Sustainability, 2021, 13(9): 4883.

[16] Chhipa H. Nanofertilizers and Nanopesticides for Agriculture [J]. Environmental chemistry letters. 2017, 15, 15-22.

[17] Dethier J J, Effenberger A. Agriculture and Development: A Brief Review of the Literature [J]. Economic systems. 2012, 36(2), 175-205.

[18] Qiu Lingyun, Wang Ming, Zhao Weidong. Improvement Research of PageRank Algorithm [J]. Software Guide, 2017, 16(02): 74-76.

[19] Huang Yan, Li Chaorong. Analysis of Web Page Importance Based on Improved PageRank Algorithm [J]. Journal of Yibin University, 2022, 22(06): 6-8+41.

[20] Su Wei. Agricultural Internet of Things platformV1.0 [Z]. Hefei Huangxin Technology Co., LTD. 2023.