RANDOM WALK FROM DIFFERENT PERSPECTIVE

Seah Choon Sen[1], Shahreen Kasim[1], Mohd Farhan Md Fudzee[1], Rusli Abdullah[2], Rodziah Atan[2]

[1]Soft Computing and Data Mining Centre, Faculty of Computer Sciences and Information Technology, Universiti Tun Hussein Onn Malaysia
[2]Faculty of Computer Science and Information Technology, Universiti Putra Malaysia
*Corresponding author email: seanseah0702@gmail.com, shahreen@uthm.edu.my, farhan@uthm.edu.my

ARTICLE DETAILS

ABSTRACT

This paper is presented to expose the different perspective of random walk. Random walk is applied in many different fields for prediction purposes and hence, it was further developed to improve the accuracy of prediction.

1. THE WORK

Random walk is one of the well-known prediction method which apply in many different field. The idea of random walk had been enhanced by many researchers to fulfil the needs of market and solve the problems in that discipline. The properties of random walk allow the researcher to tune the parameter of random walk which lead to different or constant probability of success accurate walk [1]. In the context of financial, random walk theory suggested that the prediction of changes in stock price should not only refer the past-movement and it is independent of each other. Burton Malkiel believe that the past movement or direction of the price of a stock or overall market cannot be used to predict its future movement [2]. Experiment of the truthiness of this theory had been run by Maurice Kendall. In his experiment, the price of stock goes upward over a period of times [3]. Even the probability distribution of random walk is the same, but it might bias to another side, over a period of times. The question is, how does researcher control the rate of bias towards the prediction result.

Random walk is applied in stochastic prediction of natural disasters [4]. It has been observed that a large class of random walks resulting from natural processes are not fully random. Hence, network is introduced as the complete set of states which the random walker can occupy. Networks are then classified as discrete network or continuous network. Discrete network of a random walker had limited number of nodes which are reachable by the walker on the network. While a continuous network is one where the states which are achievable by the walker are continuous in nature and therefore infinite. For such network, the probability of a random walk being at a specific location is expressed as a subset on state and probability density distribution is used to calculate it [4]. An example such network that is commonly used to describe these properties is the yeast protein interaction network (PIN) which is scale-free and also has the small-world structure [5]. Bias random walk will work if the prediction walk is fall under a predefined network.

In the context of cancer prediction, biological network is applied to optimized the accuracy of prediction. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, protein-protein interaction (PPI), artificial neural networks applied in cancer prediction and successfully increase the accuracy of prediction. Computational algorithm which combined KEGG pathway and random walk had been introduced by Liu [6]. Combination of PPI network and random walk approach had introduced by a group researcher [7]. These accurately assessing the cancer in prediction approach is introduced to identify the individual with high risk in selected diseases. Further experiment is needed to enhance the accuracy of

prediction as well as classification of selected diseases.

Significant directed random walk is introduced by applied the weight of genes on KEGG pathway while prediction process take part. Seah believe that the distance between genes will be influenced by the weight of genes and lead to the different of genes' bonding energy in applied network [8]. Vector (W) is calculated to digitize the discrimination power of walker during the prediction process. Mathematical equation is presented as below:

$$W_{t+1} = (1-r)(M)(\frac{N_1 + N_2}{2}) + rW_t$$

Higher vector value will bias the walker towards the next nodes on the applied network. Figure 1 shows the selected nodes for experimental purposes, while table 1 shows the weight of the selected genes. The resulted vector of each walk is presented in table 2.
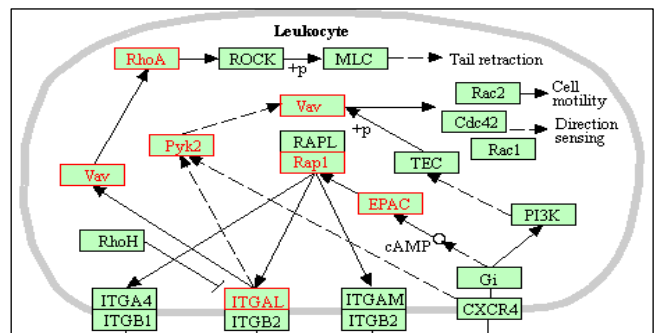

Figure 1: Gene sets that will be focusing on

Table 1: Weight of each node that focusing on

Nodes	EPAC	Rap1	ITGAL	Pyk2	Vav	RhoA
Weight	2.338914	8.47301	6.1441	3.102989	11.38365	5.149393

Table 2: Comparison result of vector from node 1 to node 6

Vector, W	Significant Directed Random Walk
W_0	0
W_1	3.243577
W_2	5.682564
W_3	5.047153
W_4	6.364853
W_5	7.505854

The fluctuation of the vector in significant directed random walk is because of the weight. Weight plays an important role to attract the other nodes. The connectivity (vector) lead to the bias walker towards the selected nodes. In short, biological network is needed to apply in random walk for cancer prediction purpose because additional information can perform as references data towards bias the walker. With this, the result of prediction can be increases due to fully utilized the biological network.

REFERENCES

[1]   Seah, C. S., Kasim, S., Mohamad, M. S. 2017. Specific Tuning Parameter for Directed Random Walk Algorithm Cancer Classification. International Journal on Advanced Science, Engineering and Information Technology, 7 (1), 176. doi:10.18517/ijaseit.7.1.1588

[2]   Malkiel, B. G. 2000. A random walk down Wall Street: including a life-cycle guide to personal investing. New York: Norton.

[3]   Kendall, M. G., Bradford Hill, A. 1953. The Analysis of Economic Time-Series-Part I: Prices. Journal of the Royal Statistical Society. A (General). Blackwell Publishing, 116 (1), 11–34.

[4]   Wijesundera, I., Halgamuge, M. N., Nanayakkara, T., Nirmalathas, T. 2016. Natural disasters, when will they reach me? Singapore: Springer.

[5]   Han, J.D.J., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J., Cusick, M.E., Roth, F.P. 2004. Evidence for dynamically organized modularity in the yeast protein interaction network. Nature, 430 (6995), 88–93.

[6]   Liu, W., Li, C., Xu, Y., Yang, H., Yao, Q., Han, J., Li, X. 2013. Topologically inferring risk-active pathways toward precise cancer classification by directed random walk. Bioinformatics, 29 (17), 2169-2177. doi:10.1093/bioinformatics/btt373

[7]   Lei, C., Ruan, J. 2012. A random walk based approach for improving protein-protein interaction network and protein complex prediction. 2012 IEEE International Conference on Bioinformatics and Biomedicine. doi:10.1109/bibm.2012.6392693

[8]   Seah, C.S., Kasim, S., Fudzee, M.F., Mohamad, M.S. 2017. A Direct Proof of Significant Directed Random Walk. IOP Conference Series: Materials Science and Engineering, 235, 012004. doi:10.1088/1757-899x/235/1/012004