RANDOM QUAD TREE AS SPATIAL DATA MINING TOOL

[1] **M. THIYAGARAJAN** , [2] **N. RAJAGOPAL**

[1] Professor, School of Computing, SASTRA University, Thanjavur, India.

[2] Senior Lecturer, Department of Mathematics, SASTRA University, Thanjavur, India.

E-mail: m _ thiyagarajan@yahoo.com, nrg1968sastra@yahoo.co.in

ABSTRACT

Among the different approaches to investigate in detail the extinction of population of rare species the Birth and Death process has given the simplest approach to settledown the solution. Here, an attempt is made to obtain the extinction probabilities and to generate a different levels of population sizes. A random quad-tree method is footforth. We give the basic concepts and results used in the subsequent sections.

Keywords: *Daubechie's Wavelet , Random Quad tree, Spatial Data Mining, Extinction probabilities.*

1. INTRODUCTION

Several attempts have been made to study a growth or Dekey of population are made through detailed study on Birth and Death process with birth and death rates of Poisson parameters. E. Seneta [8] has given the probability generating technique to solve this problem. A difficulty here is expression for probability generating function is a power of rational function of the exponential nature. Here an alternative approach is given using Spatial data mining data structure as random quad-tree for the calculation of the extinction probabilities.

2. BASIC CONCEPTS AND RESULTS

DEFINITION: 2.1 WAVELET SERIES

For any function $f \in L^2(\mathbb{R})$ we have the Wavelet series function $f(x) = \sum\sum C_{jk}\phi_{jk}$ where $C_{jk} = < f, \phi_{jk} >$ called the Wavelet Co-efficients. $\phi_{jk} = 2^{j/2}\phi(2^j x - k)\ for j,k \in \mathbb{Z}$. $\phi(x)$ is called the mother Wavelet.

The two-scale relation satisfied by Daubechies's wavelet is given by

$$\phi_3^D(x) = \frac{1+\sqrt{3}}{4}\phi_3^D(2x) + \frac{3+\sqrt{3}}{4}\phi_3^D(2x-1) + \frac{3-\sqrt{3}}{4}\phi_3^D(2x-2) + \frac{1-\sqrt{3}}{4}\phi_3^D(2x-3) \rightarrow (1)$$

$$p_0 + p_2 = \frac{1+\sqrt{3}}{4} + \frac{3-\sqrt{3}}{4} = 1$$

$$p_1 + p_3 = \frac{3+\sqrt{3}}{4} + \frac{1-\sqrt{3}}{4} = 1$$

And $p(z) = \frac{1}{2}\sum_{k=0}^{3} p_k z^k$

$$= \frac{1}{2}\left\{\frac{1+\sqrt{3}}{4} + \frac{3+\sqrt{3}}{4}z + \frac{3-\sqrt{3}}{4}z^2 + \frac{1-\sqrt{3}}{4}z^3\right\}$$

$$= \left(\frac{1+z}{2}\right)^2 \left[\frac{\left(1+\sqrt{3}\right) + \left(1-\sqrt{3}\right)z}{2}\right]$$

In fact, if the values of $\phi(1), \ldots, \phi(N_\phi - 1)$ are known, then since $\phi(k) = 0$ for all $k \le 0$ or $k \ge N_\phi$, the relations

$$\phi\left(\frac{k}{2}\right)=\sum_{l} p_l^{\phi}\phi(k-l),$$

$$\phi\left(\frac{k}{2^2}\right)=\sum_{l} p_l^{\phi}\phi\left(\frac{k}{2}-l\right),$$

etc.,

Uniquely determine all the values of $\phi(x)$ at $x=\frac{k}{2^j},\ j,k\in\mathbb{Z}$.

2.2 DETERMINATION OF THE VALUES OF $\phi(k), k\in\mathbb{Z}$

we again use the two-scale relation $\phi(x)=\sum_{k=0}^{N_\phi} p_k^{\phi}\phi(2x-k),\ p_0^{\phi},\ p_{N_\phi}^{\phi}\neq 0$, with x being an integer. That is, in matrix notation, we have $m=Mm$, where m is the column vector $m:=\left[\phi(1)...\phi(N_\phi-1)\right]^T$ and the matrix $M:=(N_\phi-1)*(N_\phi-1)$,

$$M:=\left[p_{2j-k}^{\phi}\right], 1\le j,k\le N_\phi-1 \quad \to (2)$$

with j being the row index and k the column index. Recalling that ϕ generates a partition by unity, we can determine the values of $\phi(k), k\in\mathbb{Z}$, simply by finding the eigen vector m in $m=Mm$ corresponding to the eigen value 1 and imposing the normalization condition

$$\phi(1)+...+\phi\left(N_\phi-1\right)=1 \quad \to (3)$$

DEFINITION: 2.3. QUAD TREE

One of the original data structures proposed for spatial data is that of a quad-tree. A quad tree represents a spatial object by a hierarchical decomposition of the space into quadrants(cells). This process is illustrated using the triangle. Here the triangle is shown as three shaded squares. The spatial area has been divided into two layers of quadrant divisions. The number of layers needed depends on the precision desired. Obviously, the more layers, the more overhead is required for the data structure. Each level in the quad tree corresponds to one of the hierarchical layers. Each of the four quadrants at that layer has a related pointer to a node at the next level if any of the lowest level quadrants are shaded. We label the quadrants at each level in a counterclockwise direction starting at the upper right quadrant.

3. DATA STRUCTURE IN SPATIAL DATA MINING

Because of the unique features of spatial data, there are many data structures that have been designed specifically to store or index spatial data. We briefly examine some of the more popular data structures. Many of these structures are based on extensions to conventional indexing approaches, such as B-trees or binary search trees. Non spatial database queries using traditional indexing structures, such as B-tree, access the data using an exact match query. However, spatial queries may use proximity measures based on relative locations of spatial objects. To efficiently perform these spatial queries, it is advisable that objects close in space be clustered on disk. To this end, the geographic space under consideration may be partitioned into cells based on proximity, and these cells would then be related to storage locations(blocks on disk). The corresponding data structure would be constructed based on these cells. A common technique used to represent a spatial object is by the smallest rectangle that completely contains that object, minimum bounding rectangle(MBR).

4 . OUR RESULT

The main aim of addressing the problem of computation of extinction probabilities is addressed as a problem on the spatial data mining. A spatial type of data structure to represent spatial data is a quad tree. Translating the area occupied by spatial in a major quadrant as a root and other quadrants are reached to the edges having the probabilities represented by the values of Daubechie's wavelet ϕ_3^D at dyadic points $\frac{k}{2^j}$ where $j,k\in\mathbb{Z}$. We find

$$\phi_3^D\left(\frac{k}{2^n}\right)=\frac{1}{2}\frac{\left(1+\sqrt{3}\right)^{n+1}}{4^n}, k=1, n=0,1,2,...$$

This tends to a finite limit for sufficiently large n. Thus a random quad tree represents the process of extinction of a particular spatial. This spatial data has got non-spatial features lie average environmental conditions congenial to a livelihood of a particular species.

EXAMPLE: Determine the values of $\phi_3^D(k), k \in \mathbb{Z}$, where the two-scale relation of ϕ_3^D is given by equation (1).

SOLUTION: By equation (1), we have $N_\phi = 3$ and the matrix M in equation (2) becomes

$$M = \begin{pmatrix} p_1^\phi & p_0^\phi \\ p_3^\phi & p_2^\phi \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 3+\sqrt{3} & 1+\sqrt{3} \\ 1-\sqrt{3} & 3-\sqrt{3} \end{pmatrix}$$

It is easy to see that the solution space of $m = Mm$ is $m = a\left[1+\sqrt{3} \quad 1-\sqrt{3}\right]^T, a \in \mathbb{Z}$. So, by the normalization condition (equation 3), we have $a = \frac{1}{2}$ and

$$\phi_3^D(1) = \frac{1+\sqrt{3}}{2}, \phi_3^D(2) = \frac{1-\sqrt{3}}{2}.$$

Having computed the values of $\phi(k), k \in \mathbb{Z}$, it is now very easy to compute $\phi\left(\frac{k}{2^j}\right), j, k \in \mathbb{Z}$.

5. CONCLUSION

This Random Quad tree method is better than the other approaches for the computation of extinction probabilities.

REFERENCES

[1]. C.K . Chui, “An introduction to Wavelets” , *Academic Press, INC, New York, 1992.*

[2]. M.H. Dunham, “Data Mining – Introductory and Advanced topics”, *Pearson Education, 2003*.

[3]. Bailey. N.T.J , “The Mathematical theory of Infectious Diseases & its Applications”, *Charles Griffin & Company Ltd., London, 1975*.

[4]. Bollobas. B, “Random Graphs”, *Academic Press, London, 1985.*

[5]. Hammersley. J.M. “Comparison of atom and bond percolation processes”, *J.Math. Phys. 2, PP 728 – 733.*

[6]. Kuulasmaa. K, “The Spatial general epidemic and locally dependent Random graphs”, *J. Appl. Prob. 19, PP 745 – 758, 1982.*

[7]. Kuulasmaa. K, “Locally dependent random graphs and their use in the Study of epidemic models”, *Annals of Disc. Maths, 28, PP 181 – 188, 1985.*

[8]. E. Seneta, “Functional equations and the Galton-Watson process”, *Adv. Appl. Prob. 1, 1 – 42, 1969.*

[9]. E. Seneta, “An explicit limit theorem for the critical Galton-Watson process with immigration”, *J. Roy. Stat. Soc., B, 32, 149-152, 1970.*