Advances in Computer Science and Information Technology (ACSIT)
Print ISSN: 2393-9907; Online ISSN: 2393-9915; Volume 1, Number 3; November, 2014 pp. 161-163

http://www.krishisanskriti.org/acsit.html

Stochastic Context Free Epidemic Grammar for SIR Epidemics

Mr. Senthil Athithan[1], Dr. Vidya Prasad Shukla[2] and Dr. S R Biradar[3]

[1,2]Mody University of Science and Technology, Lakshmangarh
[3]SDM College of Engineering, Dharwad

Abstract: *Human population living in a specific boundary conditions always wants to live in a disease free situation. If a disease starts to spread in the population and creates an epidemic situation, the population would then tend to spread to new location where they can be free from further infection. The method of disease spread and process of avoiding it through proper prediction has been the topic of interest of epidemiologists and computational environmentalists. Here in this work, the epidemic situation in terms of formal language and its corresponding generative grammar and the acceptance automata has been investigated.*
The need of representing the epidemic situation in terms of formal language stems from the fact that it gives more flexibility in expressing the epidemic situation rather than in numerical form. If an epidemic picture is clear then many decisions such as relocating the population or evacuation, planning vaccination strategy can be possible. The way of representing epidemic situation in one dimensional languages, then followed by two dimensional languages and the formal grammar named stochastic context free epidemic grammar has been derived for two dimensional language. The effectiveness of this grammar has been tested with the markovian chain principle which gives us the standard SIR epidemic behavior.

1. EPIDEMIC MODEL

Epidemic deals with spread of disease in an environment. There are few basic models of epidemics and they are, Susceptible-Infective-Recovered(SIR)Model (Kermack, 1927) (Porta, 2014),Susceptible-Exposed-Infective-Recovered(SEIR) Model and Susceptible-Infective-Recovered-Susceptible(SIRS) Model. The model under investigation in this work is SIR model. Susceptible population is the one which has higher probability of getting infected in to a disease. Infective population is the one which is infected and tend to infect other population which are susceptible. Recovered percentage of population are those who are recovered from the infective disease and become immune to the infection for a larger period of time

2. LANGUAGE

A one-dimensional language is the set of strings denoted by *L* and it is formally defined as $L \subseteq \sum^*$ (Ilachinski, 2002), where $\sum$ is the input alphabet and $\sum^*$ is the closure set of $\sum$. The framework for the epidemic spread is two dimensional. A two dimensional language $L_2 \subseteq \sum^{**}$ is a set of two dimensional strings in the form of a rectangular grid of cells with each cell containing a symbol from $\sum$ where $\sum^{**}$ is the set of all two dimensional strings over $\sum$.

Let us look at some of the fundamentals with regard to two dimensional languages and the assumptions to the current topic of interest, i.e. the epidemic spread.

Given a 2D-string $w_2 \in \sum^{**}$, let $r_1(w_2)$ represent the number of rows in w_2 and $c_1(w_2)$ represent the number of columns in w_2. This is to represent the population in combination of Susceptible, Infective and recovered are spread across the rows and columns in the grid. The size of the grid is termed to be $(r_1 \times c_1)$. If there are no population in the grid then it is considered to be empty string denoted by standard notation λ. Population under consideration in epidemic environment could be under some boundary conditions or it may not. Boundary conditions for example, we could set a condition that the boundary cells are immune, that means there won't be any infection spread beyond these boundaries.

3. TWO DIMENSIONAL CONTEXT FREE GRAMMAR

The epidemic two dimensional languages defined above could be generated by the two dimensional context free grammar. First normal grammar is defined and then it is modified to suit to two dimensional format.

V is the set of non-terminals or variables.

T is the set of terminals

S∈V is the special variable called Start variable. All derivation starts with S until otherwise specified.

P is set of productions, which are known to the heart of any grammar and is defined by

A➔ α where A∈V and $\alpha \in (V \cup T)^*$

Derivation process of strings which are part of the language generated by the grammar is done by starting from the symbol S and substituting the various productions in the set P in the order either by replacing the leftmost variable every time or by replacing rightmost variable. The derivation ends when we get strings of terminals.

A Two dimensional Context-Free Grammar (2DCFG) is defined by (Vh, Th, Sh, Ph) where h indicates the rules are applied horizontally across the grid. The length of the row in the grid is limited by a special variable Max(length)i. The epidemic spread context free grammar is given by:
Vh={S, I, R}
Th={s, i, r}
Sh∈Vh
where Ph is given by
{ S➔ SS / SI / IS / s
I➔ II / IR / RI / i
R ➔ RR / r }
This grammar is in Chomsky Normal Form. These production are applied parallel across each row and each row length is limited by Max(length). Number of rows are dependent on the type of geographical area that we decide.

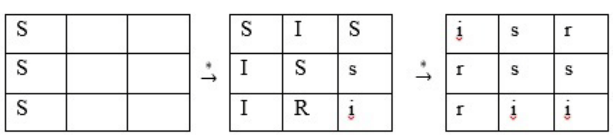

Fig. [1]:2D CFG derivation

the above Fig. depicts how the derivation of the epidemic spread context free grammar with rightmost derivation is done to obtain the final situation after getting the string of terminals which consists of the mix of susceptible, infectives and recovered population.

The only problem with the above grammar is it is ambiguous. A grammar is ambiguous if there exists at least two rightmost or two leftmost derivations of the same string. Ambiguity results in wrong prediction of the derivation and could lead to various anomalies in the results. In order to remove or reduce the ambiguity problem we can add probabilities to the productions and make the grammar as probabilistic context free grammar. (Chi, 1999)

4. PROBABILISTIC CFG

PCFG is a context free grammar with added rule probabilities under which the strings are produced is defined by (V, T, S, R, P) where V is the set of non-terminals or variables, T is the set of terminals, S is a start variable, R is the set of productions and each production would be of the form X➔ Y where X∈V and Y∈(V ∪ T)*, P is the set of probabilities by which a non-terminal is replaced during the derivation process and defined by: For each production of the form A➔x1|x2|..xi, where A ∈ V and $x1, x2 \in (V \cup T)^*$ and p1, p2..pi is the probability of probability of x1,x2..x1 getting replaced for non-terminal A, $\sum_{p=1}^{i} pi = 1$. Adding the probability to each production would give us the following production rules:

Production	Probability	Description
S➔SS	0.2	A susceptible gets added alongside of a susceptible
S➔SI	0.3	An infective gets added alongside the Susceptible
S➔IS	0.3	Susceptible gets infected and another susceptible gets added alongside.
S➔s	0.2	Susceptible remains susceptible throughout all generations.
I➔II	0.3	An infective gets added alongside the infective
I➔IR	0.2	a recovered gets added alongside an infective
I➔RI	0.3	Infective turns recovered and another infective gets added alongside
I➔i	0.2	Infective remains as infective throughout all generations
R➔RR	0.4	A recovered gets added alongside a recovered.
R➔r	0.6	Recovered remains as recovered or considered as receiving permanent immunization.

Considering a leftmost derivation for string w = "iisirs"

S➔SS➔ISS➔IISS➔iISS➔iiSS➔iisS➔iisIS➔iisIRS➔iisiRS➔iisirS➔iisirs
probability of deriving the string "iisirs"
P(w)=(0.2x0.3x0.3x0.2x0.2x0.2x0.3x0.2x0.2x0.6x0.2x0.2)= $4.1472\ x\ 10^{-8}$
Similarly the probability of all the rows of the grid is calculated. To get the picture of actual epidemic spread using the above said grammar let us apply the concept of Markov Chain.

5. MARKOV CHAIN

A Random process is called Markov Process if, the current state of the process is conditional then its future is independent of its past (Anderson, 1991).

It is a mathematical model of a random phenomenon evolving with time in a way that the past affects the future only through the present. The time could be discrete or continuous.

$P_r[X_{n+1} = x_{n+1} | X_1 \dots X_n = x_1 \dots x_n] = P_r[X_{n+1} = x_{n+1} | X_n \dots x_n$

There are three important components of Markov chain namely State space S, Initial distribution π_0 and Probability transition rule.

S is a finite or countable set of states , that is, values that the random variables Xi may take on. S = {1, 2…..N } for some finite N. This is specified by giving a matrix **P = (Pij)**. If **S** is the finite set {1…N}, say, then **P** is an N x N matrix. **"P_{ij}"** is the conditional probability, given that the chain or system is in state **i** at time **n**, say, that the chain jumps to the state **j** at time **n + 1** and it is independent of n. That is,

$$Pij = P\{X_{n+1} = j | Xn = i\}$$

The p_{ij}'s are often referred to as the transition probabilities for the Markov chain.

Now we construct the transition table for Context-Free epidemic grammar:

Table[1]: Transition Matrix for epidemic grammar

	SS	SI	IS	s	II	IR	RI	i	RR	r
S	0.2	0.3	0.3	0.2	0	0	0	0	0	0
I	0	0	0	0	0.3	0.2	0.3	0.2	0	0
R	0	0	0	0	0	0	0	0	0.4	0.6

The computation is done in an m x n cellular automata where m represents the number of rows and n represents the number of columns. In our simulation we assume that the m and n are equal. Derivation starts with Susceptible S that is the start symbol in each row and the production rules are applied either horizontally or vertically based on the following selective cellular automata neighborhood rules. In selective neighborhood a cell would be updated using the productions whenever it finds a space either in top, bottom, right or left cell whereby it updates the current cell and the blank cell as well when the right side of the production yields two nonterminals.

I	S	I		I	S	I
i	S		→	i	S	S
R	i	r		R	i	r

with the initial value of 1200 susceptible and 2 infectives and no recovered the markov chain principle has been used to find the values for 30 time steps and the following graph shows how the picture matches with the traditional SIR curve.

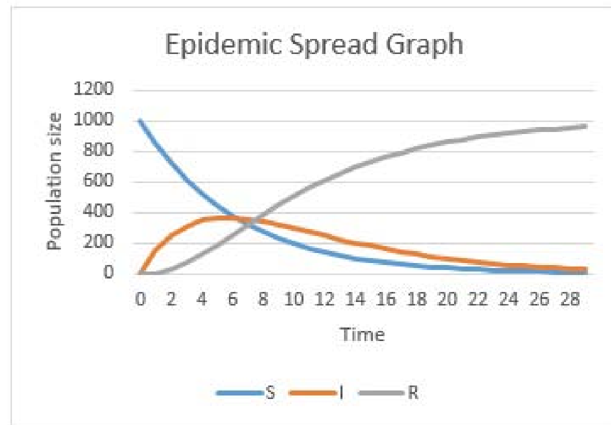

Fig. [2]: SIR curve through Markov chain values

so using the stochastic Epidemic CFG we can identify whether a particular population distribution in 2D CA can be derived or not and if it can be derived then we can identify the production rules which are the most probable ones used in that derivation.

6. CONCLUSION AND FUTURE SCOPE:

In this work a framework for modeling the epidemic spread using stochastic context free grammar is analyzed. This can be further enhanced using the context sensitive grammar instead of the context free and the resultant effect could be analyzed.

REFERENCES

[1] Chi, Z. (1999). Statistical properties of probabilistic context free grammars. *Computational Linguistics*, 131-160.

[2] Ilachinski, A. (2002). *Cellular Automata-A Discrete Universe.* World Scientific.

[3] Anderson et al. (1991). Continuous time Markov Chains. *Springer series in statistics: Probability and its application.*

[4] Kermack, M. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A*, 700-721.

[5] Porta, M. (2014). *A Dictionary of Epidemiology.* Oxford University Press.

[6] R Bonita, R. B. (2006). *Basic Epidemiology.* Geneva: World Health Organization.