A Machine Learning Model for Population Analysis among Different States in India which Influences the Socio, Demographic and Economic Needs of Society

Addepalli VN Krishna, M. Bala Murugan

Abstract: In this work Data from 2011 census is taken to identify the state which influences more in Population census among the different states identified. The data is considered from Madhya Pradesh, followed with Utter Pradesh, then to Bihar, Bengal and Orissa. Similarly other case studies are also done for Southern Indian states and North Eastern States. Genetic algorithm will be tried to find the optimal location for the given study. A fitting function is calculated for the population data of 2011 using Lagrange Interpolation technique. This fitting function is given as input to Genetic algorithm to find the optimal state which have maximum influence in the population growth among different states of India as per the Case studies done.

I. INTRODUCTION

Confrontations on Land Utilization, Regional tensions, Deterioration in environment standards, fighting for Natural resources are all the factors which are influenced by population growth. Thus systematic approaches on development became crucial in Societal, geographical and Economical issues of society. The Development and population changes are very complex issues which are inter disciplinary in nature.

Dr. Addepalli VN Krishna, Professor, CSE, Faculty of Engineering, CHRIST, Bengalurur, India. E-mail: hari_avn@rediffmail.com

Dr. M. Bala Murugan, Associate Professor, CSE, Faculty of Engineering, CHRIST, Bengalurur, India. E-mail: balamurugan.m@christuniversity.in

And also proposing any simulation approach on population studies needs a constructive methodology on fundamental concepts like Birth rate, Death rate, Exodus rate, Climatic changes and so on. The data being considered must be supportive for both Spatial and Temporal Changes.

II. LITERATURE SURVEY:

This work deals with a unified approach between Population variations and influencing factors with respect to time and population rearrangement factors across various forms of life [7]. In this work the authors developed a special model which studies the impact of Highways in space and time as parameters on population studies [8].

The authors considered Ebola Virus in Africa as a Case study in identifying patterns of spread with space and time as parameters [1]. In this work Self organizing maps are considered to study air polluted data [2]. In this work the authors worked on modeling concepts with different case studies like Traffic and House hold data analysis [3-5].

III. SCOPE & ITS SIGNIFICANCE

The scope of the work considers both internal and external factors. A model can be designed to suits the mapping between input and output values in studying the process of the mechanism. One of the significant applications of this model is Census data. Thus Census data is very important for Government organizations to plan for present and future expenditures. Planning and execution of financial resources on popular schemes which helps in betterment of societal needs can be better implemented with the help of this data. Other areas where this work can have applicability are Geography, Economics, Urban, Statistics and Operation Research, Sociology, Environmental Studies and so on.

Retrieval Number: A1066109119/2019©BEIESP
DOI: 10.35940/ijeat.A1066.109119
Journal Website: www.ijeat.org

123

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

IV. PROBLEM STATEMENT

The study discusses the population census in identified states and its influence in neighborhood states. A fitting function will be generated from the identified data which will be processed using genetic algorithm to find the most probable state which influences population among the identified set.

V. SPECIFIC OBJECTIVES

1. Updating the data on optimal influence by a state for population register.
2. Establishing the data units in the context of Economic and Social arena.
3. Make the record frame work.

VI. RESEARCH METHODOLOGY

The population at different states is considered for the given case study. In this work Genetic algorithm will be tried to find the optimal location for the given study [6]. Genetic Algorithms (GA) are non deterministic in nature with features like evolution and selection. GA's are used to simulate the living beings and are used to find the fittest of beings by using selection, mutation and cross over.

The R package supports GA tool to explore in such type of scenarios.GA tool can be used both in sequential mode and as well parallel mode. User's can define their fitting function, Population size, crossover and Mutation probabilities and expected range of outcome. The performance of the model may also be user friendly with predefined Elitism being used in GA tool.

Algorithm being used

1. The population in different states as per census 2011 is taken as input.
2. The states are considered in order from west to east to show the dependence or independence of one state from other.
3. Based on population data, a fitting function is generated using Lagrange interpolation technique.
4. By considering GA function from R tool and using the fitting function as input
   And by suitably considering Mutation and cross over probabilities, most optimal position of the sate will be calculated
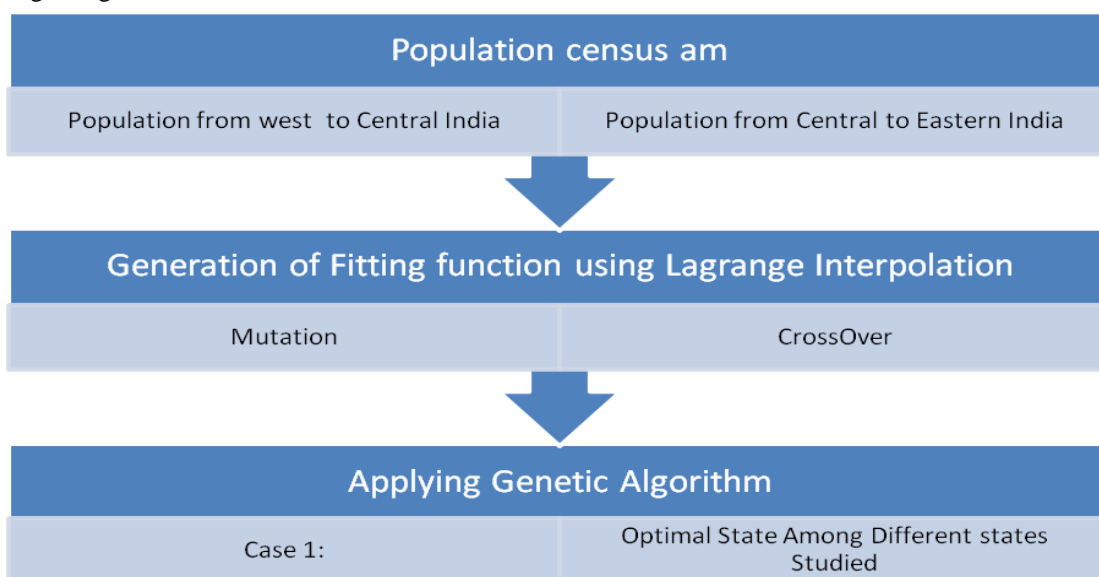
The flow of the work is shown by a Block Diagram

Case 1: Population Census is represented in value* 10 ^{6}

States Ten Year span	2011
Madhya Pradesh	72.59
Utter Pradesh	200
Bihar	103.8
Bengal	91.34
Orissa	41.95

Different cases will be studied to identify the optimal location using Genetic Algorithm. For the given problem, Data from 2011 census is taken to identify the state which influences more in Population census among the different states identified as per the table shown.

The data is considered from Madhya Pradesh, followed with Utter Pradesh, then to Bihar, Bengal and Orissa.

A fitting function is calculated for the population data of 2011 using Lagrange Interpolation technique.

$Y=f(x) = (-215/12) * x^4 + (692/3)* x^3 - (12577/12)* x^2 + (5776/3)* x -1017.$

This fitting function is given as input to Genetic algorithm to find the optimal state which have maximum influence in the population of other states as data given.

For the given problem, Considered population scope is 50, Crossover and mutation probabilities as 0.8 and 0.1 respectively, the range between 0 to 6.

Fitness function value = 210.8586

Solution =

X1= 1.727115

Results: The statistics of data generated will give the finds that state at approximate location 2 has the maximum influence on the population of the region, ie **Utter Pradesh** as per our data provided to the algorithm influences the population growth in the set of 5 states.

Case 2: Population Census is represented in value* 10 ^{6}

States Ten Year span	2011
Sikkim	0.667
Arunachal Pradesh	1.38
Nagaland	1.98
Manipur	0.272
Mizoram	1.1
Tripura	3.6
Meghalaya	2.96

Fitting function is

$Y=f(x) = (-6.45) * x^4 + (83.83)* x^3 - (376)* x^2 + (673)* x - 341.$

For the given problem, Considered population scope is 50, Crossover and mutation probabilities as 0.8 and 0.1 respectively, the range between 0 to 6.

Fitness function value = Inf

Solutions = x1= 5.868471

Results: The statistics of data generated will give the finds that state at approximate location 6 has the maximum influence on the population of the region, ie **Tripura and partly Mizoram** influences the population growth in the set

of 7 states.

Case 3: Population Census is represented in value* 10 ^{6}

States Ten Year span	2011
Kerala	33.38
Karnataka	68.62
Telangana	35.28
Andhra Pradesh	49.38
Tamilnadu	72.14

Fitting function is

$Y=f(x) = (0.019) * x^6 - (0.52) * x^5 + (5.3) * x^4 - (26.6)* x^3 + (67.9)* x^2 - (81.1)* x +35.8$

For the given problem, Considered population scope is 50, Crossover and mutation probabilities as 0.8 and 0.1 respectively, the range between 0 to 6.

Fitness function value = 74.56107

Solution =   x1= 1.65747

Results: The statistics of data generated will give the finds that state at approximate location 2 has the maximum influence on the population of the region, ie **Karnataka and partly Kerala** influences the population growth in the set of 5 states.

VII. RESULT ANALYSIS

The study identifies the larger influence of Utter Pradesh in Northern India, Influence of Tripura in North Eastern India and Karnataka in South India. This data helps the State better in decision making on Government policy decisions in improving the living standards of population.

S No.	Different Regions in India	Most Influential State
1	Northern India	Utter Pradesh
2	North Eastern India	Tripura and Mizoram
3	Southern India	Karnataka and Kerala

VIII. CONCLUSION & FUTURE WORK

As a pilot study we have considered only 3 regions ie. Northern, North Eastern Part and Southern part of India and applied Genetic algorithm to know its validation in the proposed area.

It is verified that Uttar Pradesh in North India, Tripura and Mizoram in North East India and Karnataka and Partly in Kerala are influencing the Population patterns in their respective regions. The Proposed model may be studied for more population influence of different parts of the states, which may be helpful for better policy decisions in that arena.

REFERENCES

1. Miles M.Carrol et al. ,Temporal & Spatial analysis of 2014-15 Ebola Virus outbreak in West Africa, Nature, 524, 97-101, 06 Aug 2015.

2. Lee M, Steign E, Zipf.a , Clustering and analysis air pollution data using self organizing maps, 19 AGILE conference on Geographical Information Science, Helsinki, Finland.

3. A.V.N.Krishna, A Mathematical model for estimation of Rural House Hold data through different states in India, Georgian Electronic Scientific Journal: Computer Science and Telecommunications 2009|No.1(18).

4. A.V.N.Krishna, Temporal Data Mining in Estimation of Empirical Formula for House Hold Data Analysis, IJCAI 07, Workshop on Spatial & Temporal Reasoning, IIIT H, Jan 2007.

5. A.V.N.Krishna, Spatial Analysis Model for Estimation of Population and Other Census Data in India for Forecasts in Demographic, Social and Economic Arena, Int J Adv Technol 8 (184), 2017

6. A quick tour of GA, Luca Scrucca, 07 Jun 2016

7. Guangqing Chi1 and Stephen J. Ventura, Population Change and Its Driving Factors in Rural, Suburban, and Urban Areas ofWisconsin, USA, 1970–2000, Hindawi Publishing Corporation International Journal of Population Research, Volume 2011, Article ID 856534, 14 pages, doi:10.1155/2011/856534

8. N. Ward and D. L. Brown, Placing the rural in regional development, *Regional Studies*, vol. 43, no. 10, pp. 1237–1244, 2009.

AUTHOR'S PROFILE

Dr.Addepalli VN Krishna, Professor, CSE, Faculty of Engineering, CHRIST, Bengalurur-74, hari_avn@rediffmail.com, with 25 + years of Experience and publications in Journals of repute.

Dr. M. BalaMurugan, Associate Professor, CSE, Faculty of Engineering, CHRIST, Bengalurur-74, balamurugan.m@christuniversity.in, with 13 + years of Experience and publications in Journals of repute.

Retrieval Number: A1066109119/2019©BEIESP
DOI: 10.35940/ijeat.A1066.109119
Journal Website: www.ijeat.org

126

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication