An Empirical Model of Regional Growth using Adaptive Neuro-Fuzzy Inference System

Abhishek Pandey
Department of CSE
ABES Engineering College, Ghaziabad (India)

Ashok K Sinha
Member IEEE
UST Software India Pvt. Ltd, New Delhi (India)

ABSTRACT
In India the socio-economic development of different states is spatially heterogeneous. The states can be broadly classified into three categories viz; developed, developing and underdeveloped. The development status of states falling under any one category is influenced by its socio-economic parameters. The earlier studies on regional development have analyzed the socio-economic data but no effort has been made to empirically establish the relationship among the variables in the data.. The proposed model presents an empirical model for estimating the socio-economic status of states based on Gross State Domestic Product (GSDP). The model correlating the GSDP with socio-economic parameters uses ANFIS tool for machine learning. The model so developed yields a reasonably acceptable result.

Keywords
Socio-economic parameters, learning model, ANFIS.

1. INTRODUCTION
The development of any region largely depends on socio-economic parameters such as population, power utilization, education expenditure, literacy rate, employment in organized sector, agricultural growth, poverty health, per capita income, gross state domestic product, etc. The GSDP value is the market value of final goods and services produced within a state in a year, or over a given period of time. This determines the growth status of any region/ state, whether the region/ state falls under developed, developing or under developed category. The GSDP value is aggregation of all the socio-economic indicators responsible for development of any region. While preparing regional development plans, the planners seldom use any mathematical model using various indicators. This work presents a empirical model of regional growth using adaptive neuro-fuzzy inference system **(ANFIS)** [1]**.** Using given input-output data, ANFIS construct a fuzzy inference system (FIS) whose membership functions are either adjusted by back-propagation algorithm or by hybrid learning algorithm. This adjustment allows the fuzzy system to learn from the input/ output values of the model for its modeling. ANFIS has been used because of various advantages such as (i) it refine if-then else rule for segmenting data, (ii) it does not required human expertise all the time, (iii) provides more choices of membership functions to use and (iv) it provides fast convergence time. In the proposed work, a GIS for India is developed with data obtained from various government agencies. GIS is designed to capture, store, manipulate, analyze, manage, and present all types of geographical data [2]. The model is trained with the data retrieved from the geographical information system (GIS) developed for the region, the model proposed in this paper will help the economic planners in formulating their plan policies and priorities.

2. LITERATURE REVIEW
Richard Harris [3] has presented a model which highlights the importance of both demand and supply side factors in determining regional growth. In this paper author has discussed various models which include neoclassical model, Kaldorian model, new economic geography model with importance of knowledge assets and resource based theory. In this paper author has discussed the theoretical aspects of regional growth but has not presented any computational model using the socio-economic data.

Petrakos George et al [4] developed a generalized econometric model for investigation of the determinants of regional economic growth in EU. Neo-classical growth theory, theory of endogenous growth, cumulative causation theory and new economic geography are the main components in conventional economic growth theory. In this paper author have taken most of the factors that affect regional economic growth however impacts of these factors does not shows a linear relation between development and economic regional growth since some factors are non-linear.

Natalia N. Lychkina et al [5], discussed a simulation modeling of regional social and economic development in decision support systems. In this report author have presented various methodology and technological approaches for decision support system which are based on new information technologies. This report also include the general structure of model complex for region social-economic development and its realization based on methods of system dynamics and modern technologies of simulation modeling are described.

Meir Russ et al [6], have presented a framework and set of indicators for a region which are knowledge deprived. This paper describes a systematic approach for determining the most important economic development indicators. In this paper author has also developed a three tier regional economic indicators framework to address the transition of economic regions from a manufacturing-based economy to a knowledge-based economy. However with large dataset the complexity of the model will increase.

3. HYPOTHESIS AND OBJECTIVE
For preparing regional policy plans there is a need for quantitative analysis of the socio-economic parameters such as population, education expenditure, employment in organized sector, agricultural growth rate, power utilization and gross state domestic product. The hypothesis of this work states that, "**the status (developed, developing and under developed) of any region depends upon various socio-economic indicators such as population, power utilization, employment in organized sector, education expenditure, agricultural growth rate.**" The objective is to develop a model of regional growth based on socio-economic data. The model will classify a region/state into developed, developing and underdeveloped categories based on its GSDP. This will

help in preparing development plans of a region. This will also help the planner to deciding their plan priorities beneficial for people.

4. METHODOLOGY

4.1 Data Collection

Data are identified and collected from various agencies such as World data bank, Planning Commission Government of India, ministry of Statistics and Program implementation, Education for all in India-Sarva Shiksha Abhiyan, Indiastat and Database on Indian Economy.

4.2 Development of GIS

For the development of GIS, a vector image of India has been taken and the image is registered with their longitude and latitude values of East (Long: 97.40238°E Lat: 28.01744°N), West (Long: 68.03215°E Lat: 23.71307°N), North (Long: 76.51335°E Lat: 35.88250°N) and South (Long: 76.47300°E Lat: 9.15360°N) and a point object is created for the entire region. Socio-economic data are stored at these point objects and from these point object data are retrieved.

4.3 Extraction of different socio-economic parameters using SQL queries

Socio economic parameters are extracted through SQL queries in MapInfo Professional software [7]. Queries of selecting all the relevant parameters for developing model are performed on the table and extracted data are stored in the form of text file. The data are extracted on the basis on the status of the region such as developed, developing and underdeveloped.

4.4 Development of regional model using ANFIS

ANFIS Architecture: The structure of ANFIS consists of 4 inputs and 1 output (see Figure 1). The inputs represent the different socio-economic parameters which extracted from developed GIS through SQL queries. The model is trained using sugeno-type fuzzy inference system. The sugeno system is more compact and computationally more efficient [8]. The developed model is generated with 81 FIS rules. The membership function used in this work is Gaussian constant type membership function [9].

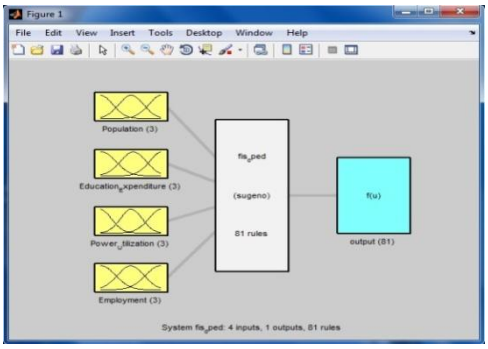

Fig 1: Adaptive Neuro-Fuzzy Inference System

Training Algorithm: The training algorithm used in this work is hybrid learning algorithm which combines the least square estimator and the gradient descent method [10]. The three (developed, developing and underdeveloped) categories are trained separately with the algorithm and the training error is calculated separately to measure the training accuracy of model.

5. IMPLEMENTATION AND RESULT

5.1 Data Analysis

The data of socio economic indicators show the statistical information about population, power utilization, employment, education, agricultural growth. These data (yearly and state wise) is shown in figure below (see Figure 2).

	Year	Popullation GR	Agriculture_Allied Sector	Education_Exp	Power_Util	Emp_Organized_Sec
2	2000	166198	0.67	5.2	25014.3	20.6
3	2001	169547	1.63	4.9	25134.2	21.7
4	2002	172944	0.14	5.3	25184.36	21.4
5	2003	176374	3.85	5.7	26659.62	20.88
6	2004	179824	-1.05	13.6	27939.25	21.1
7	2005	183282	2.34	11.1	30374.19	20.9
8	2006	186755	2.42	14.7	34543.73	21.1
9	2007	190254	3.51	14.1	37531.88	21.1
10	2008	193763	3.8	13.2	39636.82	21.2
11	2009	197271	-0.4	13.2	41250.2	21.5
12	2010	200764	4.71	16.1	44184.58	21.4
13	2011	204250	4.57	17.4	46683.3	21.5
14	2012	207739	3.47	17.6	49192.6	21.8

Fig 2: Socio-economic data

5.2 Development of GIS

The collected data are store at point object and data are visible in the Info tool box (see Figure 3).

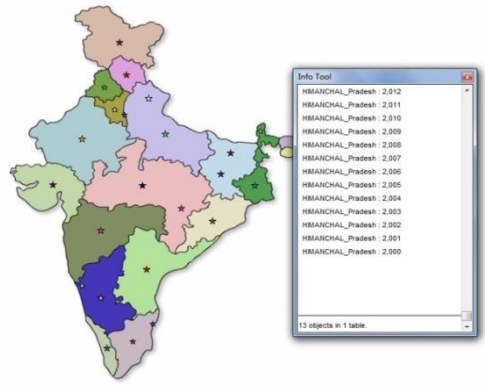

Fig 3: Data at point object

The data of particular year are stored at the point object (see Figure 4).

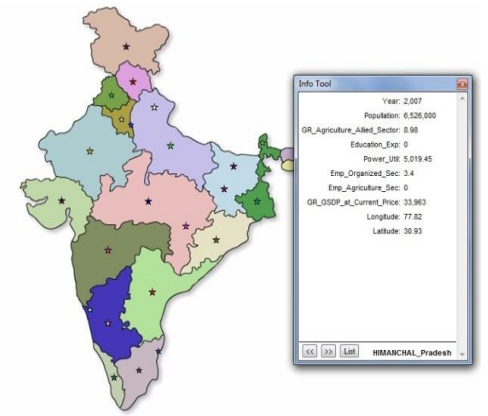

Fig 4: Data of particular year

5.3 Extraction of data from GIS

After extracting data, three data set are created for developed, developing and underdeveloped states, with statistical information of five parameters namely population (in Thousands), education expenditure (in Lakhs), power

utilization (in Gw/h), agricultural growth rate (percentage rate) and gross state domestic product (in crore). Data are normalized before training and validation (see Figure 5).

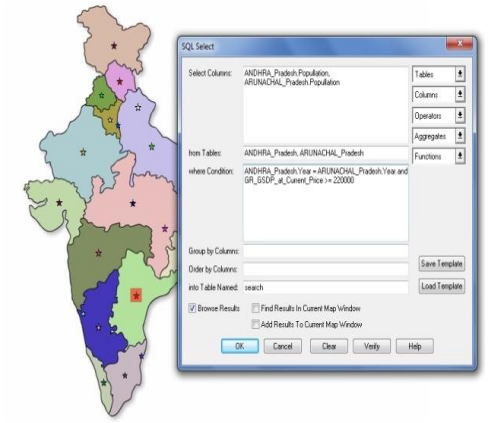

Fig 5: Data extraction through SQL queries

5.4 Development of regional growth model using ANFIS

The regional growth model is trained separately for three categories of states viz; developed, developing and under developed (see Figure 8) and there results on ANFIS tool are shown below. Training error for developed category is 3.2637e-05, 2.9022e-03 for developing category and 1.4853e-03 for under developed category.

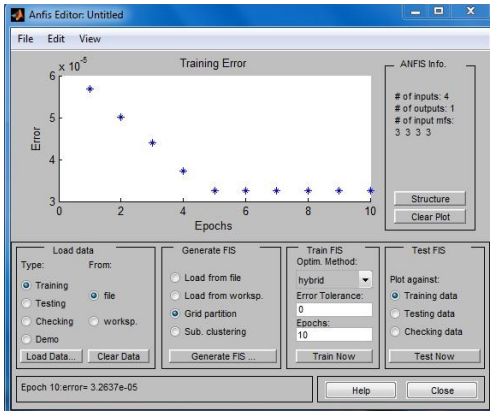

Fig 6: Trained for developed class

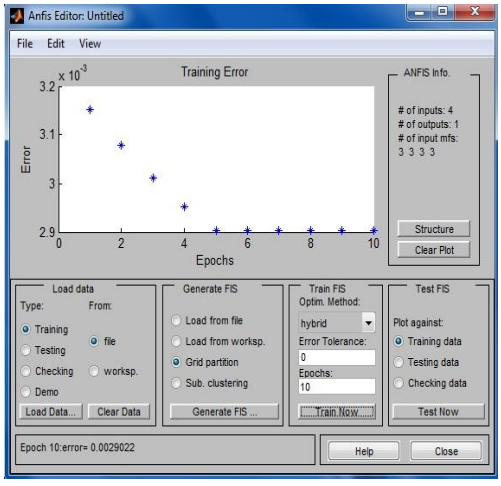

Fig 7: Trained for developing class

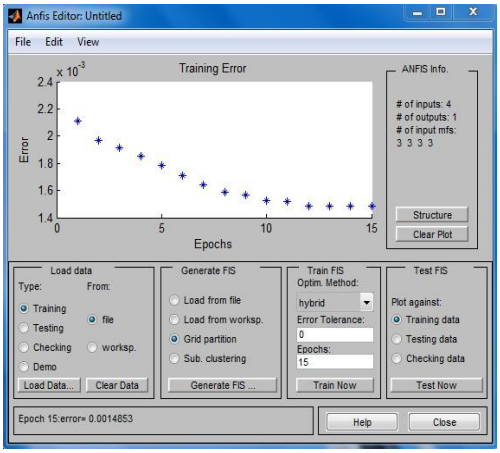

Fig 8: Trained for under developed class

6. TESTING AND VALIDATION

For validation, cross validation techniques is performed to check the accuracy of the model. Data of Bihar, Karnataka, Orissa, Tamil Nadu, Chhattisgarh, and Uttar Pradesh for year 2004, 2007, and 2009 are collected and these data are passed to all the three categories (developed, developing and underdeveloped) and validation is performed on these data and the result (see Figure 9).

States	Year	Expected Status	Actual Status	Result
Bihar	2004	Under Developed	Under Developed	Correct
	2007	Developing	Developing	Correct
	2009	Developing	Developing	Correct
Karnataka	2004	Developing	Developing	Correct
	2007	Developed	Developed	Correct
	2009	Developed	Developed	Correct
Orissa	2004	Under Developed	Under Developed	Correct
	2007	Developing	Developing	Correct
	2009	Developing	Developing	Correct
Tamil Nadu	2004	Developing	Developing	Correct
	2007	Developed	Developed	Correct
	2009	Developed	Developed	Correct
Chhattisgarh	2004	Under Developed	Under Developed	Correct
	2007	Under Developed	Under Developed	Correct
	2009	Under Developed	Under Developed	Correct
Uttar Pradesh	2004	Developing	Developing	Correct
	2007	Developed	Developed	Correct
	2009	Developed	Developed	Correct

Fig 9: Validation result of model

The result shows that the data of different region from different years are tested with all the three classes (developed, developing and underdeveloped) and result is validated satisfactorily with 88.096% accuracy.

7. CONCLUSION

The empirical work presented in this paper establishes that the development of any region depends on various socio-economic data such as population, education expenditure, employment and power utilization. The economic status is classified into three groups based on range of GSDP value viz; developed, developing and under developed. The empirical model is successfully trained and validated with the input-output data retrieved from the developed GIS for the region. The training error is 3.2637e-05 for developed, 2.9022e-03 for developing and 1.4853e-03 for under developed category. The proposed model may be used as a tool for development planning of a region and deciding its priorities. In future the model will be trained with data on more significant inputs for many more years. Cross validation will improve the result.

8. REFERENCES

[1] Jang, J.-S.R(1993)," ANFIS: adaptive-network-based fuzzy inference system", Systems, Man and Cybernetics, IEEE Transactions,Vol.23, pp665 – 685

[2] J. Star and J. Estes, "Geographic Information Systems: An Introduction". Prentice Hall, Englewood Cliffs New Jersey, 1990.

[3] Harris, Richard, 2011. Models of Regional Growth: Past, Present and Future, Wiley Journal of Economic Surveys, Vol. 25, Issue 5, pp. 913-951.

[4] Petrakos George, Kallioras Dimitris & Anagnostou Ageliki, 2007. A Generalized Model of Regional Economic Growth in the European Union. DYNREG12, Economic and Social Research Institute (ESRI).

[5] Lychkina N. Natalia and Shults Dmitriy. Simulation modeling of regions social and economic development in decision support systems.

[6] Jones K. Jeanette and Russ Mier, 2008. Regional Economic Development Indicators for a Knowledge-Based economy with Knowledge Deprivation. The Journal of Regional Analysis and Policy, Vol. 38, Issue 2, pp. 189-205.

[7] MapInfo Professional User Guide. A software documentation by Pitney Bowes.

[8] Arshdeep Kaur, Amrit Kaur. "Comparison of Mamdani-Type and Sugeno-Type Fuzzy Inference System for Air Conditioning System." 2012 International Journal of soft Computing and Engineering(IJSCE) ISSN: 2231-2307, Volume-2, issue-2.

[9] Jin Zhao, Bose, B.K. "Evaluation of membership functions for fuzzy logic controlled induction motor drive" 2002 IECON 02 [Industrial Electronics Society, IEEE 2002 28th Annual Conference of the] (Volume:1) ISBN- 0-7803-7474-6, Volume-1, page: 229 - 234

[10] Naveed Anwer, Aneela Abbas, Aneela Mazhar, Syed Hassan, 2012, "Measuring wether prediction accuracy using sugeno based adpative neuro fuzzy inference system, grid partitioning and guassmf" Computing technology and Information Management (ICCM), Vol-1, page: 214-219.