International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 12 | Dec -2016 www.irjet.net p-ISSN: 2395-0072

A Study & Survey on Rainfall Prediction And Production of Crops Using Data Mining Techniques

N. Sundaravalli [1] , Dr.A.Geetha [2]

[1] Research Scholar, PG & Research Department of Computer Science, Chikkanna Govt. Arts College, Tirupur -2. shinisuba99@yahoo.com

[2] Assistant Professor, PG & Research Department of Computer Science, Chikkanna Govt. Arts College, Tirupur -2. gee_sam@yahoo.com

---

Abstract - *Rainfall and crops production and Prediction is the most important & challenging role in the matter of all living beings in all kinds of weather happenings. It plays an essential position for countries economic growth and development in the modern world. Accurate information on rainfall is essential for the planning and management of water resources & farming. Agriculture is the main stay of the State economy in that its role is fundamentally instrumental in terms of market, factor and product contributions. Previously farmers predict the production on the foundation of assumption of experience, farmers except any assist of laptop as well as tender computing technology. There is range of algorithms for prediction purpose. Hence the present investigation primarily involves use two data mining techniques one is clustering and another one is classification. In clustering, we use the K- Means, is one of the simplest and best method for prediction work. Second we used the fuzzy classification, Finally we used the hybrid method of NeuroFuzzy with genetic algorithm. In that we conclude that, these three models & Optimization of these three K-Means, Fuzzy Logic & NeuroFuzzy with Genetic Algorithm models are useful for to predict the way of excellent & accurate result of agricultural crop production based on rainfall prediction, as it compared with the other Algorithm. This research will help the formers to know how much production will come done in next coming season and how much amount of rainfall will occur so that the formers get awareness and they can manage themselves from their heavy loss. The tool used here can be used to predict corp production based on rainfall prediction.*

Keyword: Fuzzy, K-MEANS, Prediction, Rainfall, Crop yield production.

1.INTRODUCTION

Agriculture in India has a giant history. Today, India is ranks second global in farm output. Agriculture and allied sectors like forestry and fisheries accounted for 16.6% of the GDP(Gross Domestic Product) 2009, about 50% of the complete workforce. The financial contribution of agriculture to India's GDP is gradually declining with the country's huge -based monetary growth. [1]Agriculture exercise is a type of commercial enterprise with risk. The manufacturing of Crops depends more than a few elements like on climatic, geological, organic, political and economic factors. Accurate information about the nature of historic yield of crop is vital modeling input, which is useful to farmers & Government Corporation for decision making process in organizing acceptable policies related to next production. The advances in computing and statistics storage have supplied widespread almost of data. The challenge has been to extract expertise from this raw data, Data mining that can bridge the knowledge of the statistics to the crop yield estimation. This lookup aimed to statistics mining techniques and applies them to the a number of variables consisting in the database to set up if significant relationships can be located and the usage of fuzzy logic to discover the circumstance of crops on a number stipulations of rainfalls. In this paper, an scan is made to discover the achievable of Principal aspect Analysis for broad vary forecasts of Indian summer season monsoon rainfall. Earlier this approach has drawn interest from lookup employees as it can control the elaborate and deviating options higher than the conservative statistical strategies and has effectively been utilized to a a variety of problems. Indian summer monsoon rainfall has a share of long interval common is used in this experiment.[2] The goal for the usage of these method for following reasons; to minimize the size of the data set, correspond to the information the usage of a minimal variety of factors that are capable to describe a vital proportion of the whole difference, and two to enhance the grasp of the modifications of the underlying system, by way of rendering these elements as correspond to fundamental 'modes' of local weather variability Summer precipitation in South India is related with monsoon circulation and convective activity which experiences a extensive variability over the spatial and temporal scale. In order to check out the inter-annual variability of summer time rainfall. Forecasting

frost formation on bridge-ways in Iowa is an vital yet tough problem. Frost varieties when water vapor in the air sublimates onto a floor (which takes place when the dew factor temperature of the air is greater than the floor temperature), and the floor temperature is beneath freezing. Only small quantities of moisture are wanted to cover surfaces with frost and create hazardous travel conditions. Prediction of rainfall is important in a range of contexts. The chance of rainfall is necessary for many choice makers who are touchy to the occurrence of precipitation. An accurate quantitative rainfall prediction can become aware of the manageable for heavy precipitation and possible associated flash flooding, as properly as providing facts for hydrologic interests. [3]

2. LITERATURE REVIEW

Rainfall prediction is one of the most essential and tricky job in the modern world. In general, weather and rainfall are highly non-linear and complex phenomena, which require advanced computer modeling and recreation for their accurate prediction. An Artificial Neural Network (ANN) can be used to foretell the behavior of such nonlinear systems.[4][7] Soft computing deals with approximate models where an approximation answer or result is achieved. Soft computing has three basic components, namely, Artificial Neural Network (ANN), Fuzzy logic and Genetic Algorithm. ANN is commonly used by researchers in the field of rainfall prediction.[5] Human brain is a highly complex, nonlinear, and parallel computer (information-processing system). Neural Networks are simplified models of biological neuron system. A neural network is a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experiential knowledge and making it available for use [6].The fundamental processing element of an ANN is an artificial neuron. Just like the natural neuron in human brain, it can receive inputs, process them and produce the relevant output. Neural Networks are succesful of modeling a weather forecast system.[7] Statistical indications chosen are capable of extracting the trends, which can be regarded as elements for creating the models. Statistical warning signs except coefficients of skewness and kurtosis are observed appropriate to extract the hidden patterns existing in weather data. The neural community signal processing strategy for weather forecasting is successful of yielding excellent outcomes and can be considered as an choice to usual meteorological approaches. Accurate climate forecasting performs a integral role for planning day to day activities. Neural community has been use in numerous meteorological purposes together with climate forecasting.

3. DATA COLLECTION

In this paper, The prediction of rainfall data have been collected from different by number of methods. The information process is done with three government department .like Indian Meterological department (IMD). Monsoon data set has been received from the URLwww.imd.gov.in/section/nhac/dynamic/bestpara.xls. In this dataset around 22 years cyclone data are available starting from 1983 to 2015 in which we have only consumed five years of data in the process. crops data has been collected by official website of Tamilnadu Agriculture Department .Usually, downloaded data are inconsistent and noisy in nature. We are going to preprocess and norimalize these noisy data and we get the normalized data.

4. METHODOLOGY

The followings are the common types of methods used by different researchers for rainfall predictions.

In the subject of artificial intelligence, neuron-fuzzy refers to mixtures of synthetic neural networks and fuzzy logic. Neuro-fuzzy was proposed via J. S. R. Jang. Neuro-fuzzy hybridization outcomes in a hybrid shrewd machine that synergizes these two techniques through combining the human-like reasoning style of fuzzy systems with the gaining knowledge of and connectionist shape of neural networks. [9][10] Neuro-fuzzy hybridization is broadly termed as Fuzzy Neural Network (FNN) or Neuro-Fuzzy System (NFS) in the literature. Neuro-fuzzy machine (the more famous time period is used henceforth) contains the human-like reasoning fashion of fuzzy structures via the use of fuzzy sets[11][12] and a linguistic mannequin consisting of a set of IF-THEN fuzzy rules. The major strength of Neuro-Fuzzy systems is that they are regular approximates with the potential to solicit interpretable IF-THENrules.

The power of neuro-fuzzy structures includes two contradictory necessities in fuzzy modeling: interpretability versus accuracy. In practice, one of the two residences prevails. The neuro-fuzzy in fuzzy modeling lookup subject is divided into two areas: linguistic fuzzy modeling that is centered on interpretability, normally the Mamdani model; and specific fuzzy modeling that is centered on accuracy, basically the Takagi-Sugeno-Kang (TSK) model. Although normally assumed to be the recognition of a fuzzy device through connectionist networks, this time period is also used to describe some other configurations including:

- Deriving fuzzy guidelines from trained RBF networks.
- Fuzzy logic primarily based tuning of neural network coaching parameters.
- Fuzzy good judgment criteria for increasing a community size.
- Realizing fuzzy membership characteristic through clustering algorithms in unsupervised mastering in SOMs and neural networks.
- Representing fuzzification, fuzzy inference and defuzzification through multi-layers feed-forward connectionist networks.

It should be pointed out that interpretability of the Mamdani-type neuron-fuzzy systems can be lost. To enhance the interpretability of neuro-fuzzy systems, sure measures should be taken, wherein important factors of interpretability of neuro-fuzzy structures are additionally discussed.

A current research line addresses the facts circulation mining case, the place neuro-fuzzy structures are sequentially updated with new incoming samples on demand and on-the-fly. Thereby, machine updates do no longer solely include a recursive adaptation of model parameters, but also a dynamic evolution and pruning of model aspects (neurons, rules), in order to take care of notion go with the flow and dynamically altering system behavior correctly and to hold the systems/models "up-to-date" anytime. Comprehensive surveys of a variety of evolving neuro-fuzzy structures techniques can be observed in and Pseudo outer-product-based fuzzy neural networks ("POPFNN") are a family of neuro-fuzzy systems that are primarily based on the linguistic fuzzy model Three participants of POPFNN exist in the literature:
• POPFNN-AARS(S), which is based totally on the Approximate Analogical Reasoning Scheme[5]
• POPFNN-CRI(S), which is primarily based on typically accepted fuzzy Compositional Rule of Inference[6]
• POPFNN-TVR, which is based on Truth Value Restriction
The "POPFNN" architecture is a five-layer neural network where the layers from 1 to 5 are called: enter linguistic layer, situation layer, rule layer, consequent layer, output linguistic layer. The fuzzification of the inputs and the defuzzification of the outputs are respectively carried out via the enter linguistic and output linguistic layers while the fuzzy inference is together performed via the rule, condition and outcome layers.
The mastering process of POPFNN consists of three phases:
Fuzzy membership generation.
1.Fuzzy rule identification
2. Supervised fine-tuning
various fuzzy membership generation algorithms can be used: Learning Vector Quantization (LVQ), Fuzzy Kohonen Partitioning (FKP) or Discrete Incremental Clustering (DIC). Generally, the POP algorithm and its variant Lazy POP are used to perceive the fuzzy rules. Neuro Fuzzy primarily based classification: In the present article, a novel Neuro-fuzzy classification approach is proposed. The method extracts the feature-wise information about a set of input patterns, fuzzifies its corresponding pattern values the usage of a membership characteristic (MF), and provides the degree of the memberships of individual patterns to several classes. Let us expect that we have N input patterns and M classes. Let us additionally reflect on consideration on that each sample consists of k attributes. The block graph of the proposed classification model is shown in below.

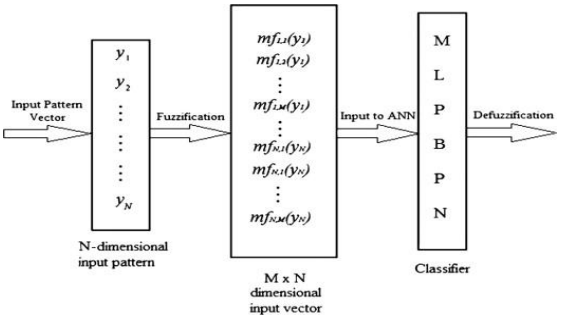

Fig 4.1 The Block graph of the classification

The method is divided into three steps which are described below.

Step 1 (Fuzzification process): In the first step, a membership matrix of order $N \times M$ is generated which consists of the degree of memberships of N different patterns to M different classes. Each element in this matrix is a membership function of the form $mf_{i,j}(y_i)$, where y_i is the i-th pattern value of input pattern vector y with $i = 1, 2, ..., N$ and $j = 1, 2, ..., M$. The MF can be defined as

$mf_{i,j}(y_i) =$ degree of membership of pattern i to class j

The input pattern vector y is thus described as $y = [y_1, y_2, ..., y_N]^T$

Where 'T' denotes the matrix transpose operation. For fuzzification, we have used the generalized bell-shaped MF which depends upon three different parameters a, b, and c as given by the equation

$$mf(y; a, b, c) = \frac{1}{1 + \left|\frac{y-c}{a}\right|^{2b}}$$

Each of these configuration parameters has some special significance: parameter c determines the center of the MF; a denotes half-width; and b (together with a) controls the slopes at the different crossover points. Below shows a typical

bell-shaped MF. By modifying the values of a, b, and c; we will obtain our desired MF which provides more flexibility for classification.

Fuzzy with genetic algorithm based classification:

Genetic fuzzy systems are fuzzy systems constructed by using genetic algorithms or genetic programming, which mimic the process of natural evolution, to identify its structure and parameter.

When it comes to automatically identifying and building a fuzzy system, given the high degree of nonlinearity of the output, traditional linear optimization tools have several limitations. Therefore, in the framework of soft computing, genetic algorithms (GAs) [13] and genetic programming (GP)[14] methods have been used successfully to identify structure and parameters of fuzzy systems.

Fuzzy systems are fundamental methodologies to represent and process linguistic information, with mechanisms to deal with uncertainty and imprecision. For instance, the task of modeling a driver parking a car involves greater difficulty in writing down a concise mathematical model as the description becomes more detailed. However, the level of difficulty is not so much using simple linguistic rules, which are they fuzzy. With such remarkable attributes, fuzzy systems have been widely and successfully applied to control, classification and modeling problems (Mamdani, 1974) (Klir and Yuan, 1995) (Pedrycz and Gomide, 1998).

Although simplistic in its design, the identication of a fuzzy system is a rather complex task that comprises the identication of (a) the input and output variables, (b) the rule base (knowledge base), (c) the membership functions and (d) the mapping parameters.

Usually the rule base consists of several IF-THEN rules, linking input(s) and output(s). A simple rule of a fuzzy controller could be:

IF (TEMPERATURE = HOT) THEN (COOLING = HIGH)

The numerical impact/meaning of this rule depends on how the membership functions of HOT and HIGH are shaped and defined.

The construction and identification of a fuzzy system can be divided into (a) the structure and (b) the parameter identification of a fuzzy system.

The structure of a fuzzy system is expressed by the input and output variables and the rule base, while the parameters of a fuzzy system are the rule parameters (defining the membership functions, the aggregation operator and the implication function) and the mapping parameters related to the mapping of a crisp set to a fuzzy set, and vice versa. (Bastian, 2000).

Much work has been done to develop or adapt methodologies that are capable of automatically identifying a fuzzy system from numerical data. Particularly in the framework of soft computing, significant methodologies have been proposed with the objective of building fuzzy systems by means of genetic algorithms (GAs) or genetic programming (GP).[15][16]

5. RESULT

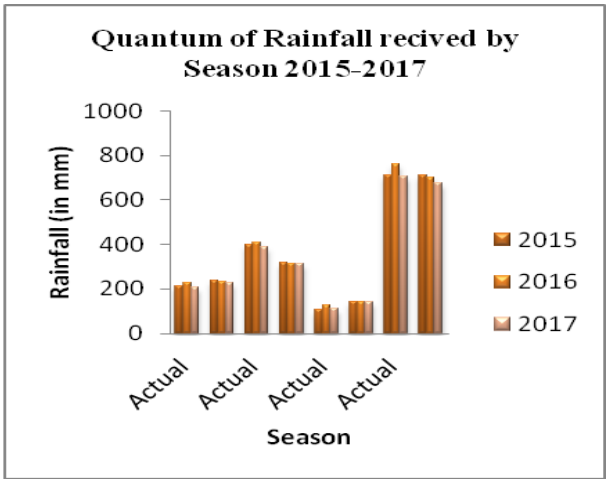

Chart 5.1 Seasonal Rainfall Prediction

The given Table 4.1 show the seasonal rainfall during the year 2015- 2017

Table 5.1 Seasonal Rainfall

Year	South - West Monsoon		North - West Monsoon		Hot - Weather Season		Whole Year	
	A	N	A	N	A	N	A	N
2015	214	234	399	315	107	143	708.6	710.6
2016	228	229	409	314	124	142	762.6	702.9
2017	207	225	388	310	112	140	707.4	675.3

A-Actual N-Normal

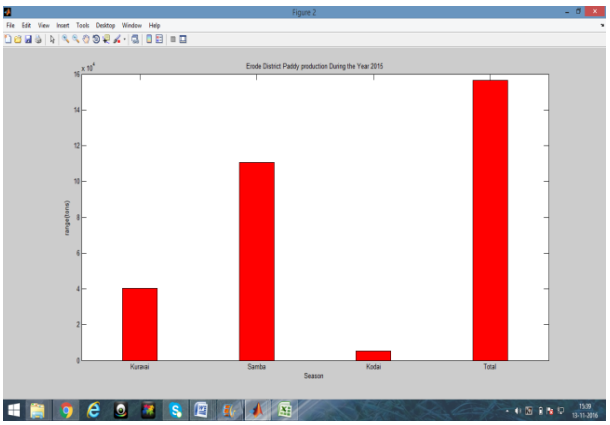

Fig 5.1 Season wise Paddy Production

The above result have been obtained from using neuro fuzzy_GA in the MATLAB platform. MATLAB is a programming language developed by MathWorks. It started out as a matrix programming language where linear algebra programming was simple. It can be run both under interactive sessions and as a batch job. MATLAB (matrix laboratory) is a fourth-generation high-level programming language and interactive environment for numerical computation, visualization and programming. It allows matrix manipulations; plotting of functions and data; implementation of algorithms; creation of user interfaces; interfacing with programs written in other languages, including C, C++, Java, and FORTRAN; analyze data; develop algorithms; and create models and applications.

It has numerous built-in commands and math functions that help you in mathematical calculations, generating plots, and performing numerical methods.

MATLAB is widely used as a computational tool in science and engineering encompassing the fields of physics, chemistry, math and all engineering streams.

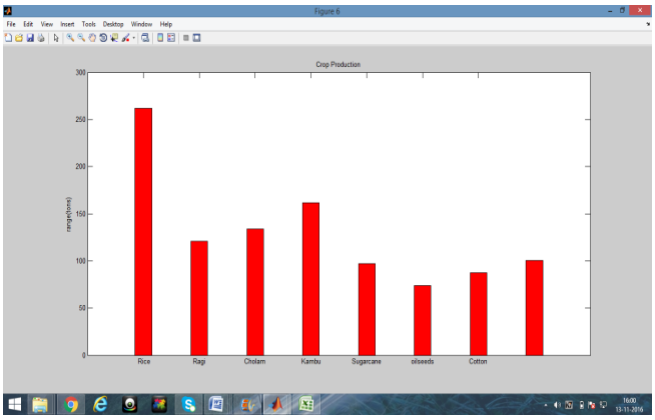

Fig 5.2 Over All Food grains Production

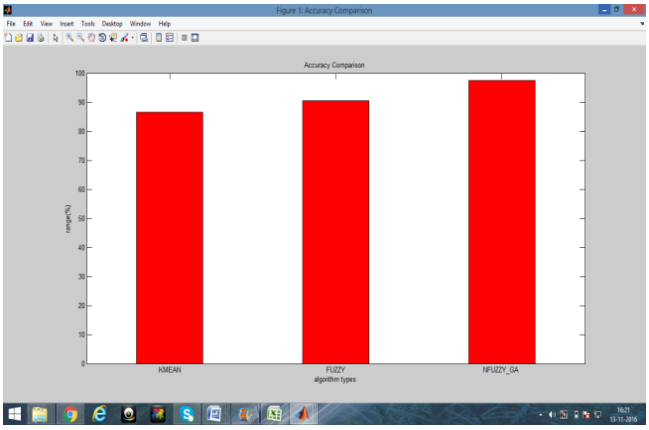

Fig 5.3 Accuracy Comparsion of Algorithm

From the observation of results from NFUSSY_GA algorithm give more accuracy than Fuzzy and KMEAN algorithm are show in the above figure 5.3

The given Table 5.2 show the accuracy and time period of KMEAN,FUZZY and NFUZZY_GA algorithm. From the above result By doing NFUZZY_GA we are obtaining the most accurate 97.4% results and less time period 1.8 minites compared to the other algorithm

Table 5.2 Comparsion of Acuuracy & Time Period

S.NO	Algorithms	Accuracy	Time Period
1	KMEAN	86.6	3.3
2	FUZZY logic	92.8	2.5
3	NEURO FUZZY_GA	97.4	1.8

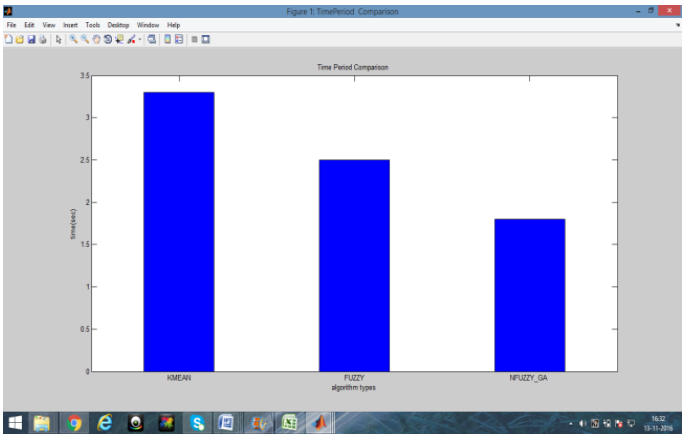

Fig 5.4 Time Period Comparsion

6. CONCLUSION

This lookup will help the farmers to recognize how plenty production will be done in next coming season and how a great deal quantity of rainfall will appear so that farmers can get awareness and they can control themselves from their heavy loss. The device used here can be used to predict coproduction of any place definitely uploading the facts in accordance to users.

7. FUTURE WORK

With the improvement of laptop technologies, mainly those without any premises or people subjective, fuzzy good judgment can be applied in many areas. In this paper some fuzzy logic were adopted in order to estimate crop production with existing data and their use in K-means& Fuzzy logic. Still there are some technique have now not but applied to agriculture problem, such as Actual climate situation for genuine date to know farmers to harvest and yield or graduation of seeds and many more applied sciences to be employed for discovering important information from agricultural-related like soil identification, pest manage and etc.

REFERENCES

[1] Dr. K. Jothi Sivagnanam,"State agriculture profile of TamilNadu,2012" Agro-Economic Research Centre, University of Madras, Chennai, Tamil Nadu, India.

[2] P. Singh and B. Borah,"Indian summer monsoon rainfall prediction using artificial neural network", stochastic Environmental Research and Risk Assessment,2013,1436-3240.

[3] A.Geetha & Dr,G.M Nasira, "Rainfall Prediction Using Logistic Regression Techniques",CiiT International Journal of Atrificial Intelligence system & Machine Learning, ISSN 0974-9667,vol 6,No.07,PP.246-250,Aug 2014

[4] A.Geetha & Dr.G.M Nasira,"Artificial Neural Networks",Applications in Weather Forecasting-using Rapidmines",Published in International Journal of Computational Intelligence and Informatics,vol.4:No.2,October-December 2014.

[5] S. Chattopadhyay and Manojit Chattopadhyay,:A Soft Computing technique in rainfall Forecasting", International Conferences on IT,HIT,March 19-21,2007.

[6] S. Haykin, 1999: Neural Networks- A Comprehensive Foundation. Addison Wesley Longman.

[7] Dr.S. Santhosh Baboo and I. Khadar Shareef,"An efficient Weather Forecasting Model using Artificial Neural Network", International Journal of Environmental Science and Development,Vol. 1,No.4,October 2010.

[8] Takagi ,T.,and M. Sugeno,"Fuzzy identification of system and its application to modeling and control",IEEE Transactions on Systems,Man,and Cybermetics,15:116-132,1985.

[9] Jang, J.-S.R., Sun,C.-T. Neuro-fuzzy modelling and control. Proceedings IEEE 83(3), 378-406,1995.

[10] Fuzzy logic tool box,user guide,The Math works.

[11] J.Yen and R. Langari,"Fuzzy Logic:Intelligence,Control,and Information",Prentice Hall,1999.

[12] Jang, J.-S.R.,ANFIS: adaptive network based fuzzy inference system.IEEE Transaction on Systems, Man and Cybernetics 23 (3),665-683,1993.

[13] Shipra Banik,Mohammed Anwer,A.F.M, Khodadad Khan,Rifat Ara Rouf and Farah Habib Chanchary, "Forecasting Bangladeshi monsoon rainfall using neural network and genetic algorithm approaches", International Technology Management Review, volume 2 Number 1, 2009

[14] J.C. Bezdek, R.J. Hathaway, Optimization of fuzzy clustering criteria using genetic algorithms, Proc. 1st IEEE Conf. on Evolutionary Computation (EC-IEEE' 94), vol. 2, 1994, pp. 589 –594.

[15] F. Herrera, M. Lozano, J.L. Verdegay, Tuning fuzzy controllers by genetic algorithms, Internat. J. Approx. Reason. 12 (1995) 299–315.

[16] F. Herrera, J.L. Verdegay (Eds.), Genetic Algorithms & Soft Computing, Springer, New York, USA, 1996. [7] J.H. Holland, Outline for a logical theory of adaptive systems.