Optimised Subtractive Clustering for Neuro-Fuzzy Models

JUUSO RANTALA, HANNU KOIVISTO
Institute of Automation and Control
Tampere University of Technology
P.O. Box 692, FIN-33101 Tampere
FINLAND

Abstract: - This paper presents results obtained when developing more efficient clustering methods for neuro-fuzzy model identification. Nelder-Mead optimisation is applied for fine tuning subtractive clustering based rule selection parameters. The performance is tested with various data sets against each other and earlier works done. The proposed method seems to produce more accurate models with fewer rules.

Key-Words: - Optimization, Subtractive clustering, Neuro-Fuzzy model, Nelder-Mead, Identification

1 Introduction

Fuzzy system identification is a well-established field and various methods are available ranging from membership function refinement to automatic rule generation, c.f. [10],[4]. The rule extraction phase is still not straightforward but more like state-of-the-art. Problems encountered during the identification phase consider typically: 1) how many rules are enough, 2) which input variables should be taken into account.

This paper addresses both problems and presents new results, which seems to produce more accurate models with fewer rules. The basic idea is to fine tune subtractive clustering based rule generation by optimising candidate rule selection parameters. Nelder-Mead (NM) is applied in this paper although other direct search methods should give similar results. Also genetic algorithms (GA) were tested but NM seems to be a reasonable compromise between complexity and efficiency.

The method is presented in Chapter 2 and its behaviour is analysed Chapter 3 which also contains comparisons with other methods.

2 Method descriptions

2.1 Subtractive clustering

Subtractive clustering algorithm [5,6] uses data points as candidates for cluster centers. Data matrix with *n* data points $\{x_1,...,x_n\}$ in *M* dimensional space includes also output and it is normalized within hypercube.

Density Measure for all data points is calculated first. For x_i it is defined as :

$$D_i = \sum_{j=1}^{n} e\left(- \frac{\| x_i - x_j \|^2}{(r_a / 2)^2} \right) \quad (1)$$

if data point has many neighbouring points it will has high density measure. r_a defines the circle where neighbouring data points lies and points outside of this circle has only little effect. Data point with highest value is selected as first cluster center.

Let x_{c1} be the selected point and D_{c1} its density measure. r_b is then used as radius which defines the neighbourhood where the density reduction is done.

$$D_i = D_i - D_{c1} e\left(- \frac{\| x_i - x_{c1} \|^2}{(r_b / 2)^2} \right) \quad (2)$$

After reduction is done data point with highest density measure is selected as cluster x_{c2}. Next we use cluster x_{c2} for calculating density reductions and this phase is repeated until $D_{ck} > AD_1$, (where A is positive constant). If $D_{ck} \le RD_1$, (where R is smaller positive constant than A), the process ends. But if $RD_1 \le D_{ck} \le AD_1$, we have to calculate

$$\frac{d_{\min}}{r_a} + \frac{D_{ck}}{D_1} \ge 1 \quad (3)$$

(where $d_{\min}$ is shortest distance between x_k and previously found cluster centers) and if it is true x_k is selected as cluster center x_{ck} and process continues. If it is not true we set potential of x_k to

zero and select the data point with next highest density to test $RD_1 \le D_{ck} \le AD_1$. This phase continues until $\frac{d_{\min}}{r_a} + \frac{D_{ck}}{D_1} \ge 1$, is not true. Last step is to calculate sigma values for the found clusters.

$$s = \left(\frac{(r_a * (\max(X) - \min(X)))}{\sqrt{8}} \right) \quad (4)$$

(where X is data matrix) . Those values are needed with membership function parameters calculation.

Proper selection of rule selection parameters requires some trial-and-error. Major goal of this study is to analyse the effect of those adjustable clustering parameters for rule generation.

2.2 Cluster estimation with least squares estimator and ANFIS

Cluster estimation with a least squares estimation algorithm is the method introduced by Chiu [5] and implemented in Matlab Fuzzy Toolbox [6]. It is one-pass method to take input-output data and generate first order Sugeno fuzzy inference system. We will later introduce Optimized subtractive clustering which contains this Chiu's [5] method. ANFIS (Adative Neuro-Fuzzy Inference System) is a neural network point of view for Sugeno type fuzzy model, proposed by Jang [1,4]. Where the initial structure is done by Cluster estimation with least squares estimator and hybrid learning algorithm combines the gradient descent and least-squares estimator method for the optimal parameters search.

The difference between network structures in Cluster estimation with least squares estimator and ANFIS is in membership functions.

Cluster estimation with least squares estimator:

$$gaussian\ (x;c,s) = e^{-\frac{1}{2}\left(\frac{x-c}{s}\right)^2} \quad (5)$$

ANFIS:

$$bell\ (x;a,b,c) = \frac{1}{1 + \left|\frac{x-c}{a}\right|^{2b}} \quad (6)$$

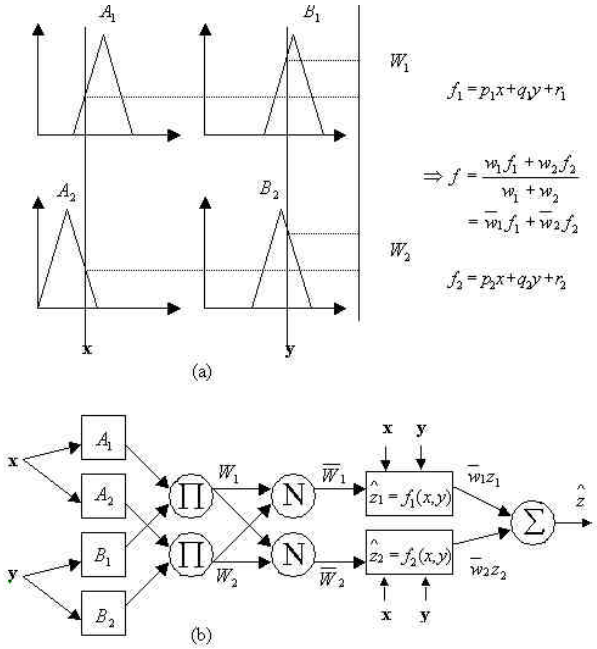

Fig.1. (a) First order Sugeno fuzzy model, (b) Corresponding network architecture. [3]

2.4 Optimized clustering

The Subtractive clustering algorithm [4,5] is difficult to adjust by conjectural way. This is especially true when some input selection method like Genetic Algorithm (GA) is used to reduce input dimensions. Compromise between time and sufficient accuracy led to decision where Subtractive clustering [4,5] parameters r_a and r_b were taken to optimize, (A=0.5 and R=0.15 were kept fixed).

Model was constructed by using training data and test RMSE error was the criterion of model quality. With test RMSE criterion we assumed that both training and test data are from same region which was not true in all cases (Box-Jenkins gas furnace data).

Root Mean Squared Error:

$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$, where y_i is actual output and $\hat{y}_i$ is desired output. (7)

In Fig. 2, is presented r_a and r_b parameters effect for cost function with MPG data set. It is clear that gradient based methods will have difficulties with sharp edges where gradient does not exist. Those sharp edges particularly appears when number of rules changes. This fact led to decision that optimization should rather do by direct search method.

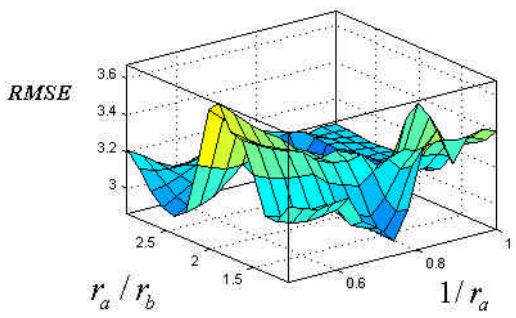

Fig. 2, Cost function surface where MPG data is evaluated with different r_a and r_b combinations.

Nelder-Mead simplex algorithm [7,9], first published in 1965, has become one of the most used direct search method for nonlinear unconstrained minimization tasks. The presentation of algorithm and its convergence properties in low dimensions can be found from reference [7,9].

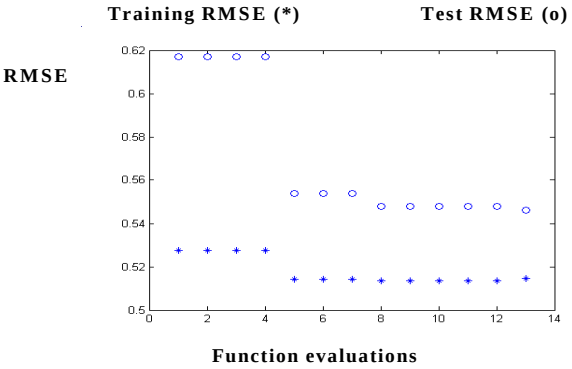

Fig. 3, Nelder-Mead test and training error convergence with Suburban Commuting data set.

Table 1. Values used to create Fig.3.

Function evaluations	Training RMSE	Test RMSE	r_a	r_b
0	0.5276	0.6170	2	1.6
1	0.5276	0.6170	2	1.6
2	0.5276	0.6170	2	1.6
3	0.5276	0.6170	2	1.6
4	0.5142	0.554	2.2222	1.6537
5	0.5142	0.554	2.2222	1.6537
6	0.5142	0.554	2.2222	1.6537
7	0.5135	0.5478	2.5	1.8181
8	0.5135	0.5478	2.5	1.8181
9	0.5135	0.5478	2.5	1.8181
10	0.5135	0.5478	2.5	1.8181
11	0.5135	0.5478	2.5	1.8181
12	0.5147	0.546	2.6667	1.855
.				
20	0.5152	0.5457	2.717	1.8306

Nelder-Mead [7,9] was chosen because of its fast convergence in low dimensions. Although local minimum search is not absolutely correct in theoretical sense but some global minimum search for example GA should use. However those global search methods are much slower than this and in this case several parameter combinations provide near the global minimum solutions, which are quite easy to find.

When this method is used together with GA based input selection, the Nelder-Mead optimization is done with one initial value where r_a = 2 and r_b = 1.6. After the best input combination is found the multi-patch optimization is used from four initial points.

3 Results with data sets

3.1 Suburban Commuting

This benchmark considers relationship between the number of automobile trips generated from an area and the area's demographics [6]. Demographic and trip data are from 100 traffic analysis zones in New Castle County, Delaware. Five demographic factors are considered: population, number of dwelling units, vehicle ownership, median household income, and total employment. Hence the model has five input variables and one output variable.

We used first 75 data points as training set and last 25 points as test set. All inputs were used.

Table 2. Results with Suburban commuting data set.

Model	Training RMSE	Test RMSE	Rules
Optimized subtractive clustering	0.5147	0.546	4
Optimized subtractive clustering+ANFIS training	0.512	0.533	4
Subtratctive Clustering+ANFIS training [6]	0.3407	0.5827	

The major improvement with this data set follows from effort of Optimized subtractive clustering and ANFIS training can achieve only little improvement after it.

3.2 Automobile MPG prediction

This benchmark considers Automobile MPG (miles per gallon) prediction [3] where several input variables are used to predict one continuous output variable. Input variables contains information about automobiles:

Multi-valued discrete	
Displacement:	Continuous
Horsepower:	Continuous
Weight:	Continuous
Acceleration:	Continuous
Multi-valued discrete	

We have 392 data points, training and test sets are randomly selected. Both contains 196 data points.

Table 3. MPG data set with inputs: horsepower, weight and model-year.

Model	Training RMSE	Test RMSE	Rules
Optimized subtractive clustering	2.5522	2.7154	4
Optimized subtractive clustering+ANFIS training	2.5274	2.714	4

With MPG data set we also tested GA based input selection. The best combination we found is horsepower, weight and model-year. This combination with Optimized subtractive clustering led to much better results than Jang [3,4] with two inputs. There is hardly any difference between Optimized subtractive clustering and ANFIS training results.

Table 4. MPG data set with inputs: weight and model-year.

Model	Training RMSE	Test RMSE	Rules
Optimized subtractive clustering	2.6808	2.8637	6
Optimized subtractive clustering+ANFIS training	2.6808	2.8637	6
ANFIS training [3,4].		2.98	

With weight and model-year we could also achieve better results than Jang [3,4]. ANFIS training couldn't improve model after Optimized subtractive clustering.

3.3 Box-Jenkins gas furnace data

This benchmark is a widely used for system identification [3,8]. There are originally 296 data points {*y(t),u(t)*}, from *t=1…296. y(t)* is the output CO2 concentration and *u(t)* is the input gas flow rate. Here we are trying to predict *y(t)* based on {*y(t-1)… y(t-4), u(t-1)…u(t-6)*}. This reduces the number of effective data points to 290.

Most methods find that the best set of input variables for predicting *y(t)* is {*y(t-1),u(t-4)*}. Sugeno and Yasukawa [8] has found that the best set of input variables for predicting *y(t)* is {*y(t-1), u(t-4), u(t-3)*}. We found as a best set for predicting *y(t)* is {*y(t-1) y(t-2) y(t-3) u(t-2)*}.

Training set is first 145 data points and test set is last 145 points.

Table 5, Results with Box-Jenkins data set.

Model	Training RMSE	Test RMSE	Rules
Optimized subtractive clustering	0.1350	0.3576	5
Optimized subtractive clustering+ANFIS training	0.1350	0.3576	5

Numerical information from Jang's [3] results was not available. We found that ANFIS training couldn't improve any after Optimized subtractive clustering.

3.4 Mackey-Glass time-series

Mackey-Glass (MG) time-delay differential equation [2,4] is a well known and widely used benchmark problem in the neural network and fuzzy modeling research communities.

$$\dot{x}(t) = \frac{0.2x(t-t)}{1+x^{10}(t-t)} - 0.1x(t) \qquad (8)$$

This time-series is chaotic, it will not converge or diverge and the trajectory is highly sensitive to initial conditions.

Fourth-order Runge-Kutta method is used to find numerical solution to equation (8). We assume *x(0) = 1.2*, $t = 17$, and *x(t)* = 0 for *t* < 0.
Inputs:{*x(t–18) x(t–12) x(t–6) x(t)*}.
Output: *x(t+6)*.
Data from *t=118…1117*, first 500 data points is training data and last 500 test data.

Table 6. Results with Mackey-Glass data set.

Model	Training RMSE	Test RMSE	Rules
Optimized subtractive clustering	0.0029	0.0034	59
Optimized subtractive clustering+ANFIS training	0.000615	0.0011	59
Optimized subtractive clustering	0.01	0.01	14
Optimized subtractive clustering+ANFIS training	0.0014	0.0013	14
ANFIS training [2]	0.0016	0.0015	16

It is clear that the proposed approach with gaussian membership function is not ideal for a highly nonlinear model like this. ANFIS with gbell membership function has more expression power and better performance. This can clearly be seen from results in table 6.

This true only in noiseless situation and it is not so straightforward when some noise is added to identification data (more practical situation). See table 7.

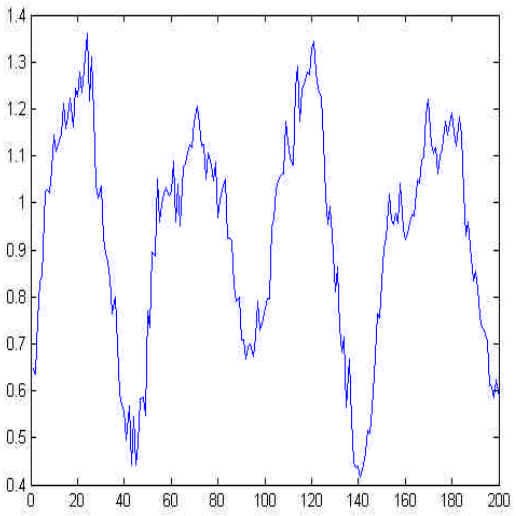

Fig. 4. First 200 points from Mackey-Glass data with zero-mean normal distributed noise (s = 0.03).

Table 7. Results with Mackey-Glass + zero-mean, normal distributed noise (s =0.03).

Model	Training RMSE	Test RMSE	Rules
Optimized subtractive clustering	0.0435	0.0482	10
Optimized subtractive clustering+ANFIS training	0.0395	0.0458	10

3.6 Prediction model

As a generalization test of Optimized subtractive clustering and ANFIS training we used iteratively 90 steps ahead predicted Mackey-Glass time-series. This kind of test is definitive way to check whether the time-series model is correctly found or not. Results of test can be found from Fig. 5.

At first we took 10 delayed values {*x(t), …, x(t-9)*} from Mackey-Glass time-series as input candidates to predict *x(t+1)*. First 500 data points were used as training set and next 600 as test set. Last 90 data points we wanted to predict was not used during the identification process and model was kept fixed during the prediction. As a best set of inputs to predict *x(t+1)* we found {*x(t), x(t-1), x(t-3), x(t-8), x(t-9)*}.

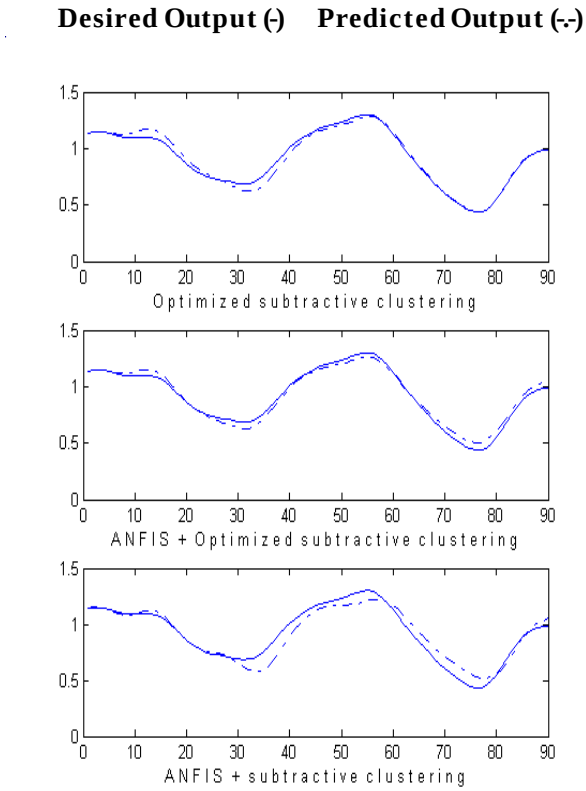

Fig. 5. Recursively ninety steps ahead computed Mackey-Glass time-series.

Models where Optimized subtractive clustering is used can achieve better results than ANFIS training after subtractive clustering with default settings. The main reason is number of rules which is smaller with default settings.

4 Conclusion

In this paper we have presented quick and efficient iterative way to construct close to optimal first-order Sugeno fuzzy model. By using this Optimized subtractive clustering we have achieved a good results without ANFIS training and with it. Also the way it works in input selection is very promising. Method is tested by using nonlinear regression and time-series data sets. Test of generalization capability can be found from part where iterative way calculated 90 steps ahead prediction for Mackey-Glass time-series is done.

References:

[1] Jang, J.-S.R., Adaptive-Network-Based Fuzzy Inference System, *IEEE Transactions on, Systems, Man and Cybernetics*, Vol. 23 Issue: 3, May-June 1993, pp. 665–685.

[2] Jang, J.-S.R.; Sun, C.-T., Predicting Chaotic Time-Series with Fuzzy If-Then Rules, *Second IEEE International Conference on, Fuzzy Systems 1993*, Vol.2, pp. 1079-1084,

[3] Jang, J.-S.R., Input Selection for ANFIS Learning, *Proceedings of the Fifth IEEE International Conference on Fuzzy Systems, 1996*, Vol.2, pp. 1493-1499

[4] Jang J.-S.R, Sun C.-T., E. Mizutani, *Neuro-Fuzzy and Soft Computing, A Computational Approach to Learning and Machine Intelligence*, Prentice Hall, 1997.

[5] Chiu, S.L., A cluster estimation method with extension to fuzzy model identification, *IEEE World Congress on Computational Intelligence, Proceedings of the Third IEEE Conference on 1994*, Vol.2, pp. 1240-1245

[6] *The Fuzzy Logic Toolbox, User's Guide version 2*, The Math Works Inc.

[7] Lagarias, J.C., J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions," *SIAM Journal of Optimization*, Vol. 9 No. 1, 1998, pp.112-147.

[8] Sugeno, M., Yasukawa, T., A fuzzy-logic-based approach to qualitative modelling, *IEEE Transactions on Fuzzy Systems*, Vol.1 Issue: 1, Feb. 1993, pp. 7

[9] Margaret H. Wright, AT&T Bell Laboratories, Direct-Search Methods: Once Scorned, Now Respectable, *Biennial Dundee Conference on Numerical Analysis*, Dundee, Scotland, 1995.

[10] H.Hellendoor, D.Driankov, *Fuzzy Model Identification, Selected Approaches*, Springer-Verlag, 1997.