

Stochastic Search in Data-Based Modelling of Dynamic Systems

Shane Butler[†] and John Ringwood*

*Department of Electronic Engineering
NUI Maynooth, Ireland*

Abstract — This paper examines the application of stochastic search techniques for the solution of two typical problems in modelling nonlinear systems using a multi-modelling approach: interpolation function determination and linear model structure determination. Two candidate stochastic search techniques are employed, genetic algorithms and swarm intelligence, which show different advantages for each of the problems considered.

Keywords — Stochastic search, genetic algorithms, swarm intelligence, multiple models

I INTRODUCTION

The issue of nonlinear data-based modelling using multiple linear models has been comprehensively dealt with in the literature. In particular, the pioneering work by Sugeno [1] had paved the way for much that has followed, as exemplified by [2, 3]. While the issue of interpolating or switching between individual linear models must be treated with care in the control situation to ensure that global closed-loop stability is achieved [4], the main concern in the modelling and/or forecasting application is the determination of an interpolating or switching function which maximises the model accuracy with respect to some criterion. Unfortunately, such a problem is not usually convex, which renders it unamenable to traditional optimisation methods, such as gradient descent. Once a candidate interpolation function has been determined, however, the determination of the parameters of the linear models is relatively straightforward and can be evaluated using a variety of linear identification techniques [5].

The only outstanding problem then remaining is the determination of the structure of the set of linear models, with either a uniform structure across all models, or individual model structures, being chosen. Traditionally, this issue is resolved in one of two ways:

1. Correlation-based techniques are used to determine cut-offs for regressors (seasonal and non-seasonal), or
2. The performance of a variety of model structures is examined with respect to some criterion (which may include a complexity cost) and the optimal structure selected.

Typically, the criterion employed in method 2. above is the mean square error across the training or validation data, with supplemental complexity cost added in via criteria such as the Minimum Description Length (MDL) or Akaike's Information Theoretic Criterion [5]. However, the models dealt with in such a manner are normally restricted to those with contiguous regressors, as exemplified in the excellent MatlabTM System Identification Toolbox. Method 1. above using, for example, the Box-Jenkins methodology [6], provides for seasonal models, but the interpretation of the partial correlation functions can be a little subjective.

The paper addresses the dual problems of interpolation function optimisation and linear model structure determination. The interpolation function is chosen to be a set of overlapping (fuzzy) functions (in the spirit of TSK-type models) while the objective is to provide an arbitrary combination of regressor components in the synthesis of the linear model(s).

Both problem domains represent non-convex optimisation problems and two stochastic optimisation techniques, both of which are inspired by natural phenomena, are selected as candidate solution methodologies. In particular, optimisation techniques based on natural selection [7] and swarm intelligence [8] are shown to be suited to such problems and have solution structures which match the problem domains.

II PROBLEM DOMAINS

The two problems considered in this study both relate to the general nonlinear dynamical modelling problem, posed in the form of a TSK problem.

a) *Determination of linear model structure*

The problem of linear model structure determination may be articulated by considering the rational (z) transfer function in (1):

$$G(z) = \frac{b_0 + b_1 z^{-1} + \dots + b_m z^{-m}}{1 + a_1 z^{-1} + \dots + a_n z^{-n}} z^{-d} \quad (1)$$

The most common structure determination procedure involves the evaluation of suitable values of m , n and d . This makes the assumption that all of the terms from z^0 to z^{-m} (for the input regressor) and z^{-1} to z^{-n} (for the output regressor) are involved in the determination of the system output. An alternative formulation, popular in time series modelling, is to make the assumption that the regressor (in particular the output regressor) is structured into the Seasonal AutoRegressive Integrated SARI(p,D)(P,DD) form, given as:

$$\Phi_p(b)\Phi_P(B^L)\nabla_L^{DD}\nabla^D Y_t = a_t \quad (2)$$

where a_t is a white noise sequence and B the delay operator. Φ_p and Φ_P are the non-seasonal and seasonal regressors respectively, of order p and P . ∇_L^{DD} and ∇^D are seasonal and non-seasonal differencing operators, with:

$$\nabla_L^{DD}\nabla^D(B) = (1 - B^L)^{DD}(1 - B)^D \quad (3)$$

contributing to the ‘Integrated’ designation. Focussing on the regressor terms:

$$\begin{aligned} \Phi_p(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \\ \Phi_P(B^L) &= 1 - \phi_{1,L} B^L - \phi_{2,L} B^{2L} - \dots - \phi_{p,L} B^{pL} \end{aligned}$$

the regressor product therefore (assuming $L > p$) contains the following terms:

$$\begin{aligned} \Phi_p(b)\Phi_P(B^L) &= 1 - \phi_1 B - \dots - \phi_p B^p \\ &- \phi_{1,L} B^L + \phi_1 \phi_{1,L} B^{L+1} + \dots + \phi_p \phi_{1,L} B^{L+p} - \dots \\ &- \phi_{p,L} B^{pL} + \phi_1 \phi_{p,L} B^{pL+1} + \dots + \phi_p \phi_{p,L} B^{pL+p} \end{aligned}$$

Clearly, the structure above does not require contiguous regressor terms, but nevertheless the set of available structures is restricted.

The proposal is to determine an algorithm which will allow an arbitrary structure to be found, according to an appropriate criterion, with specification only on the maximum (numerator or denominator) regressor length. This can, conceptually at least, be attempted using the form in (1) by examining the variance of each identified parameter values relative to its nominal value. However, considerable subjective judgement is required and no information is available on the effect of complete omission of parameter ‘bins’.

Appropriate criteria to employ in structure determination include MDL and AIC, given as:

$$J_{MDL} = V(1 + \log(N)) \frac{np}{N} \quad (4)$$

$$J_{AIC} = V(1 + \frac{2NP}{N}) \quad (5)$$

where V is the loss function:

$$V = \frac{1}{N} \sum_{k=1}^N (y_k - \hat{y}_k)^2 \quad (6)$$

with \hat{y}_k being the predicted system output. In the case of the SARI model, the Box-Jenkins [6] procedure utilises inspection of the Sample AutoCorrelation Function (SACF) and the Sample Partial AutoCorrelation Function (SPACF). However, interpretation of the SACF and SPACF for order determination (p and P) requires some experience and is somewhat subjective. The season length, L , must be specified *a priori*.

Overall, the structure determination problem for a linear system model represents a non-convex optimisation problem and will be examined for the transfer function:

$$G(z) = \frac{z^{-3} + 0.5z^{-5} + 0.3z^{-10}}{1 + 0.2z^{-1} + 0.6z^{-10}} \quad (7)$$

for a range of signal-to-noise ratios.

b) *Interpolation function synthesis*

In the TSK formulation, interpolation is performed across a set of linear models, such as those described in Section a). This can be represented in Fig.1, with the overall model determination pro-

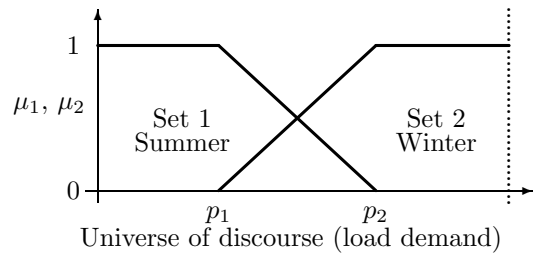


Fig. 1: Fuzzy sets for data partitioning

cedure outlined in Fig.2. The issue of consequent function order was dealt with in Section a), and the determination of the consequence parameters represents a linear least-squares problem, so the focus here is on the determination of the optimal data partitioning parameters, p_1 and p_2 . This problem has been examined before, with a benchmark problem described in [9], developing a TSK model to predict weekly electricity load. A SARI consequent model (as in (2)) is used since the

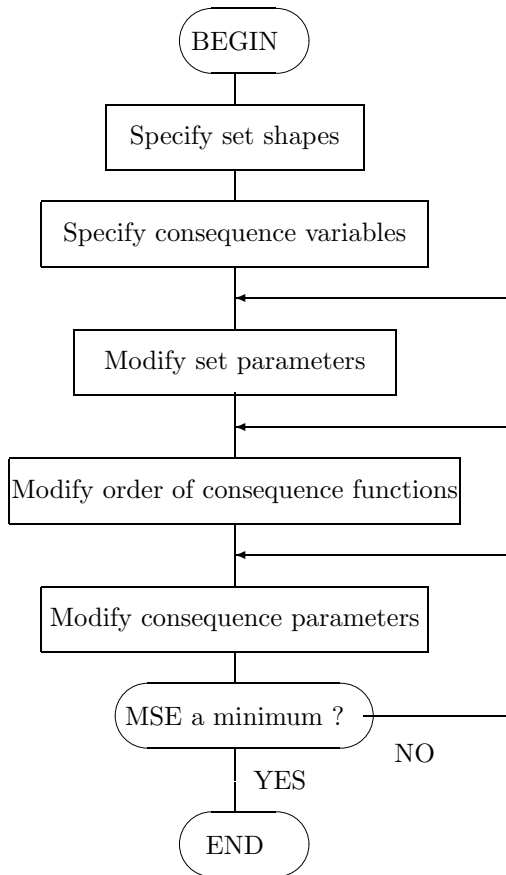


Fig. 2: Model determination procedure

data is seasonal (see Fig.3 and two fuzzy sets are employed, observing the dichotomy between summer and winter consumption behaviour. The universe of discourse, in this case, is the consumption in week $(k - 1)$, where the consumption in week k is being forecast. The determination of the fuzzy partitioning parameters (the p_i), for the load modelling problem, represents a challenging non-convex optimisation problem, as evidenced by the performance surface shown in Fig.4.

III STOCHASTIC OPTIMISATION TECHNIQUES

Clearly, both problems presented in Section II represent non-convex optimisation problems and are not solvable using traditional gradient optimisation. Stochastic search techniques maintain a ‘family’ of solutions which learn from each other, with other mechanisms guaranteeing a diversity of solutions, which ensures that a global optimum is found.

a) Genetic algorithms

Genetic algorithms (GAs) are well described elsewhere [7] so, for brevity, the attention here will focus on the salient points in relation to the problem domain.

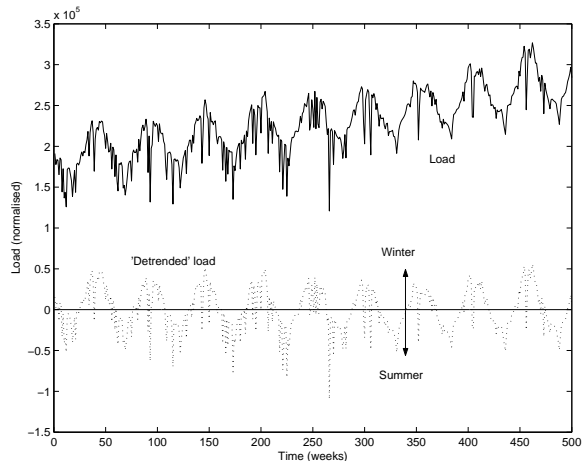


Fig. 3: Weekly electricity load data

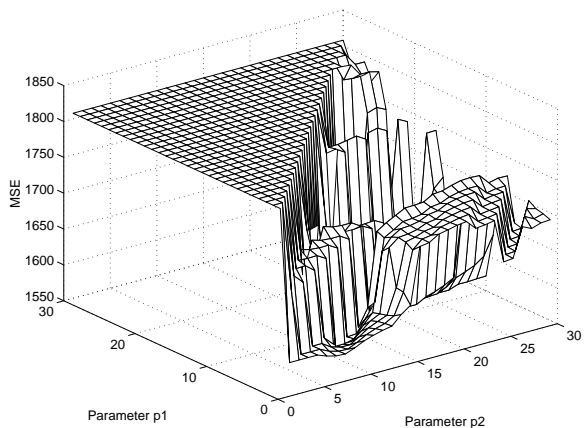


Fig. 4: Performance surface for p_1 and p_2

GAs, by maintaining a ‘population’ of solutions and utilising a number of features which allow combination of solutions (e.g. crossover) and extension of the solution set (e.g. mutation) can provide good global search capability. Chromosomes can be coded using binary or real numbers, depending on whether the problem is discrete or continuous. In some cases GAs are combined with gradient search to improve the accuracy of the final solution for continuous problems. GAs can solve a wide range of discrete problems, with a compact representation using binary-values chromosomes. Indeed, they present themselves as an ideal solution mechanism for the problem described in Section II(a) where each binary ‘bit’ represents inclusion (or exclusion) of a regressor parameter. The chromosome length represents the maximum regressor order. For example, the transfer function in (7) would be coded as in Fig.5. Chromosomes can be coded using real or binary numbers [10]. However, given the stochastic nature of the search, no guarantee of achieving the exact global minimum can be had, even for real coding.



Fig. 5: Chromosome representation for model structure

b) *Swarm intelligence*

The foraging behaviour of ants can be artificially mimicked to solve combinatorial optimisation problems, as evidenced in the early work of Deneubourg [11]. Ant colony intelligence is harnessed by recreating the behaviour of ants in laying *pheromone* trails which allow inter-ant communication and indicate the optimal paths to follow. Pheromone also evaporates over time, leading to dynamic optimisation.

A particular problem, that has received considerable attention from the ‘ant colony’ community is the travelling salesman problem (TSP) [12], which has close parallels with foraging-like activity. However, application of this paradigm to non-TSP problems is less than straightforward, even to other optimisation problems where the solution domain is discrete.

The authors are aware of only one publication applying swarm intelligence to problems where the solution domain is continuous [13], though the problem is articulated in [14]. In essence, the approach here consists of the following procedure:

1. A starting point (the ‘nest’) in the solution space is chosen (can be arbitrary)
2. An equally-spaced set of direction vectors is established
3. Ants are assigned probabilistically to search directions
4. Ants now start a distance X along chosen directions from starting point
5. The direction vectors are weighted with values depending on the concentration of pheromone, indicating the success of ants that have chosen to explore each vector
6. Ants are re-assigned to directions based on pheromone and roulette wheel (probabilistic) selection
7. Ants move a random distance (up to max. Y) in a random direction
8. Overall directions from nest are updated, along with pheromone weighting
9. Y is decreased (with each iteration)
10. Go to 5.

$S/N(dB)$	∞	32	28	20	<20
AIC	0	0.0338	$\frac{z^{-13}}{z^{-3}}$	X	X
MDL	0	0.0338	0.0586	0.1234	$\frac{z^{-4}, z^{-6}}{z^{-3}, z^{-8}, z^{-11}, z^{-12}}$

Table 1: Performance of GA-based structure identification using AIC and MDL criteria

An example of this type of operation is shown in Fig.6. Note that X and Y (and the decrease factor for Y) are parameters of the swarm algorithm (along with the initial ‘nest’ location).

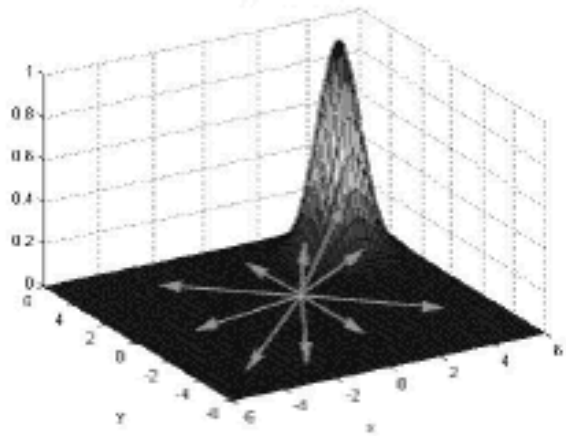


Fig. 6: Example of swarm optimisation [13]

IV RESULTS

a) *Determination of linear model structure*

The system in (7) was examined under a range of signal-to-noise (S/N) conditions, with the maximum order of numerator and denominator regressor terms both set at 15. A data sequence of 1000 points was utilised, with 500 for training (determination of parameters) and 500 for validation (determination of structure). The first 200 training points were discarded to remove the initial condition response. The input sequence chosen was zero mean, resulting in a zero mean output sequence. Additive noise was normally distributed and zero mean.

Table 1 gives an overview of the achieved performance, based on training and validation data records of 500 points each, utilising both AIC and MDL as criteria. Where a number is included, it denotes the MSE for a model with correct structure, while ‘X’ denotes incorrect structure. The included terms (powers of z) are the extra ones added, at the largest S/N where the algorithm fails to identify the correct structure. Clearly, the MDL is superior as a complexity-weighted criterion for this particular application. By way of comparison, Table 2 gives the parameter estimates and their standard deviations (SDs) for the maximal regres-

sor size (order 15) for a low-noise case ($S/N = 32\text{dB}$). Note that a number of significant ($SD < |\text{coeff}|$) erroneous parameters are generated. The (erroneous) selected model order, using both AIC and MDL, is the maximal order of 15, using the validation data for model structure verification, as in the case of the GA-based algorithm.

b) Interpolation function synthesis

For the individual seasonal models, the season length is clearly 52 (weeks), as verified by Fig.3, with the seasonal and non-seasonal regressor orders determined as $P = 2$ and $p = 3$ respectively. 12 initial direction vectors were established with 24 ants assigned randomly to these directions, using roulette wheel selection. The initial distance from the nest, X , was set at $1/4$ of the overall parameter range (60,000), while the maximum subsequent random movement was $1/2$ the overall parameter range, with a reduction factor of 0.85 per iteration. The actual random distances were taken from a normal distribution, with the appropriate maximum value. Fig.7 shows the evolution of the mean squared error (MSE) for the total model, with the adaptation of the fuzzy parameters (shown in Fig.1) performed using the continuous swarm algorithm as described in the preceding section. The

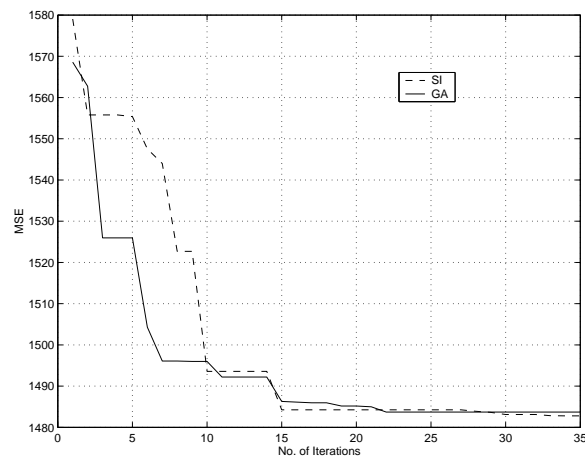


Fig. 7: Comparative performance of genetic and swarm paradigms

comparative performance of a GA-based adaptation, as reported in [9], is also shown as a benchmark. In general, the swarm algorithm performs at least as well (with a minor improvement) compared to the GA-based technique. The iteration count for the GA-based algorithm corresponds directly to generation count, allowing some comparison to be made on computational complexity. The convergence rates for both algorithms are comparable, but it should be noted that the GA had a population of 70, while only 24 ants were utilised.

A typical prediction from the TSK model is shown in Fig.8, with the performance of a single

linear model shown for comparison. The Mean Absolute Percentage Errors (MAPEs) for TSK and linear models are 2.5% and 3.9% respectively, indicating significant improvement for the nonlinear model.

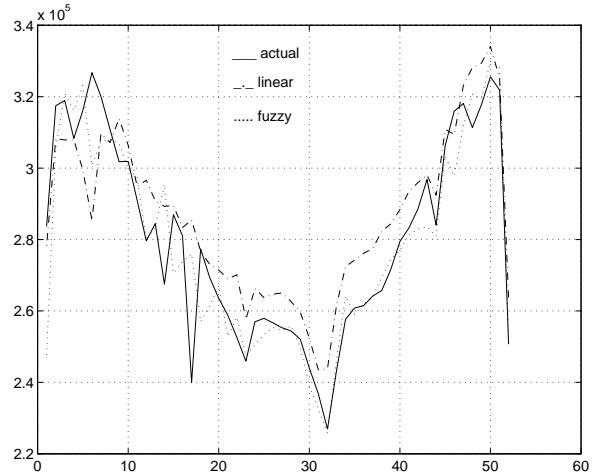


Fig. 8: Load prediction performance of TSK and linear models

V CONCLUSIONS

The paper has examined the application of search techniques based on evolutionary and natural intelligence to the nonlinear dynamic modelling (multi-model) problem. Problems which are discrete in the solution space (model structure) and continuous (determination of interpolating parameters) arise, both of which are non-convex and cannot be solved using traditional gradient techniques. The GA-based solution to the model structure issue exploits the natural representation of chromosomes, while the swarm algorithm can outperform, if only modestly, the GA algorithm in the continuous problem. One explanation for this is that, for continuous problems, the swarm algorithm can be considered as a type of 'parallel' gradient algorithm, where future search directions are informed by the previous solutions (with some random selection), while the 'adaptation gain' is progressively reduced (consistent with the gradient $\rightarrow 0$ as the minimum is reached) with progressive iteration. Due to their relatively recent development, artificial swarm-type algorithms have been rarely applied and this paper suggests that there is considerable scope for them in areas of modelling and control, given that their accuracy is no worse than GAs and they have (for the particular problem considered here) a significant computational advantage.

Term	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7
Coef.	1.0000	0.0127	-0.0013	-0.1265	0.0244	-0.0729	-0.0390	-0.0270
SD	0.0000	0.0587	0.0590	0.0593	0.0595	0.0587	0.0185	0.0152
Term	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}
Coef.	0.0052	-0.0020	0.5785	-0.1062	0.0302	-0.0827	0.0229	-0.0574
SD	0.0152	0.0150	0.0146	0.0368	0.0373	0.0373	0.0375	0.0370
Term	b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7
Coef.	0.0062	0.0100	-0.0019	0.9913	-0.1936	0.5313	-0.2288	0.0980
SD	0.0111	0.0111	0.0111	0.0111	0.0592	0.0606	0.0681	0.0694
Term	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}	b_{14}	b_{15}
Coef.	-0.1358	0.0014	0.2398	-0.0581	-0.0114	-0.0635	0.0291	-0.0221
SD	0.0687	0.0359	0.0351	0.0247	0.0249	0.0251	0.0250	0.0249

Table 2: Results from Least Squares

REFERENCES

- [1] T. Takagi and M. Sugeno. "Fuzzy identification of systems and its applications to modelling and control". *IEEE Trans on Systems, Man and Cyber.*, **15**, 117-132, 1985.
- [2] T.A. Johansen and B.A. Foss. "Construction NARMAX models using ARMAX models". *Int. J. of Control*, **58**, 1125-1153, 1993.
- [3] (-) "Special issue on multiple model approaches to modelling and control. *Int. J. of Control*, **72**, 7/8, 1999.
- [4] D. Liberzon and A.S. Morse. "Basic problems in stability and design of switched systems". *IEEE Control Systems Magazine*, **19**, 5, 59-70, 1999.
- [5] L. Ljung *System Identification: Theory for the User (2nd Ed.)*, Prentice Hall, 1999.
- [6] G.E.P. Box, G.M. Jenkins and G.C. Reinsel. *Time Series Analysis: Forecasting and Control (3rd Ed.)*, Prentice Hall, 1994.
- [7] D.E. Goldberg *Genetic Algorithms in Search, Optimisation and Machine Learning*, Addison-Wesley, 1989.
- [8] E. Bonabeau, M. Dorigo and G. Theraulaz *Swarm Intelligence: From Natural to Artificial Systems*, Oxford University Press, 1999.
- [9] J.V. Ringwood "Optimisation of fuzzy electricity forecasting models using genetic algorithms". *Proc. 5th Euro. Congress on Intelligent Techniques and Soft Computing (EUFIT)*, Aachen, Germany, Sept. 1997, Vol.3, 2457-2461.
- [10] C.Z. Janikow and Z. Michalewicz "An experimental comparison of binary and floating point representations in genetic algorithms". *Proc. 4th Intl. Conf. on Genetic Algorithms*, 1991, 31-36.
- [11] J.-L. Deneubourg, S. Goss, J.M. Pasteels, D. Fresneau and J.-P. Lachaud "Self-organising mechanisms in ant societies (II): Learning in foraging and division of labour" *Experientia Suppl.*, **54**, 1987, 177-196.
- [12] M. Dorigo and L.M. Gambardella "Ant colony system: a cooperative approach to the travelling salesman problem". *IEEE Trans. Evol. Comp.*, **1**, 1997, 53-66.
- [13] L. Kuhn "Ant colony optimisation for continuous spaces". *BE Thesis, University of Queensland*, Oct. 2002.
- [14] G. Bilchev and I.C. Parmee "The ant colony metaphor for searching continuous design spaces". *Proc. AISP Wkshp. on Evol. Comp.*, 1995, 25-39.