Evolving Similarity Coefficients for the Solution of Cellular Manufacturing Problems

C. Dimopoulos
University of Sheffield,
Department of Automatic Control & Systems Engineering,
Mappin Street, S1 3JD, Sheffield, U.K.
cop97cd@sheffield.ac.uk

N. Mort
University of Sheffield,,
Department of Automatic Control & Systems Engineering,
Mappin Street, S1 3JD, Sheffield, U.K.
n.mort@sheffield.ac.uk

Abstract- **The cell formation problem is a classic manufacturing optimisation problem associated with the implementation of a cellular manufacturing system. A variety of hierarchical clustering procedures have been proposed for the solution of this problem. Essential for the operation of a clustering procedure is the determination of a form of similarity between the objects that are going to be grouped. In this paper we employ a Genetic Programming algorithm for the evolution of new similarity coefficients for the solution of simple cell formation problems. Evolved coefficients are tested against the well-known Jaccard's similarity coefficient on a large number of problems taken from the literature..**

1 Introduction

The elimination of unnecessary costs is a significant issue in the design of a modern manufacturing system. The concept of Group Technology (GT) (Mitrovanov, 1966), (Burbidge, 1975) has been proposed as a guide for the implementation of cost-effective manufacturing systems that do not require heavy investment in new equipment. GT states that the grouping of objects performing similar tasks within a corporate structure can yield substantial benefits. On the shop-floor level this proposition is translated as the grouping of machines into a number of cells that process a dedicated number of part families. The term 'cellular manufacturing' is traditionally used to describe such a system. It has been reported (Singh, 1993) that cellular manufacturing systems reduce set-up costs, work-in-progress inventories and throughput times while improving the quality of products and the job satisfaction of workers..

The implementation of a cellular manufacturing system is a multi-input, multi-stage procedure that requires human expertise. A particular stage of this procedure, the cell formation problem, has attracted considerable research attention during the last thirty years due to its combinatorial nature. Reviews of alternative solution methodologies that have been proposed over the years are presented in Selim *et al.* (1998) and Dimopoulos and Mort (1999).

Evolutionary Algorithms (EA's) have been extensively used for the solution of manufacturing optimisation problems (Dimopoulos and Zalzala, 1999). In this paper we report a novel solution methodology based on Genetic Programming (Koza, 1992) and a classic hierarchical clustering procedure.

The remainder of the paper is organised as follows: In section 2 Jaccard's similarity coefficient and its relevance to the solution of the cell formation problem is described. The GP-SLCA framework for the evolution of new similarity coefficients is presented in section 3. The performance of evolved similarity coefficients is illustrated in section 4. The conclusions of this report are drawn in section 5.

2 Jaccard's similarity coefficient and the cell formation problem

Finding the optimal grouping of machines is not an easy task since the number of possible combinations is astronomical even for small problem instances (Lee and Garcia-Diaz, 1993). The use of similarity coefficients focuses the search for an optimal solution to regions of the search space where it is more likely to be found. More specifically, Jaccard's similarity coefficient, as used by McAuley (1972) in his pioneering paper, defines the similarity between a pair of machines as follows:

$$S_{ij} = \frac{a_{ij}}{a_{ij} + b_{ij} + c_{ij}}$$

where:

S_{ij} : similarity between machines i and j
a_{ij} : number of parts processed by both machines i and j
b_{ij} : number of parts processed by machine i but not by machine j
c_{ij} : number of parts processed by machine j but not by machine i

Similarity information can be used by a clustering procedure for the creation of potential machine-cells. In McAuley's paper the Single Linkage Cluster Analysis

(SLCA) (Sneath, 1957) procedure was employed for the construction of a 'dendrogram' of potential solutions which were distinguished by different levels of similarity (an illustration of McAuley's procedure is presented in the Appendix of this paper, Tables A-E). After the machine-cells were created, part families could be easily identified. While Jaccard's coefficient is quite simple and efficient, its application is independent of the optimisation objective and relies on the existence of natural groupings of machines. In this paper we employ a GP platform for the evolution of a similarity coefficient that will be competitive to Jaccard's similarity coefficient on the solution of simple cell formation problems.

For the representation of the cell formation problem we employ the traditional form of the machine-component (m/c) matrix. The m/c matrix uses binary notation to indicate the existence of a part operation on a particular machine in the plant. An example of an m/c matrix is given in Figure 1.

	p1	p2	p3	p4	p5
m1	0	1	1	0	1
m2	1	0	0	1	0
m3	0	1	1	0	1
m4	1	0	0	1	0

Figure 1: An example of an m/c matrix

The diagonalised m/c matrix is normally used for the illustration of the created machine cells and the corresponding part families (Figure 2).

	p4	p1	p5	p3	p2
m2	1	1	0	0	0
m4	1	1	0	0	0
m1	0	0	1	1	1
m3	0	0	1	1	1

Figure 2: The diagonalised m/c matrix

3 The proposed methodology for the evolution of new similarity coefficients

The framework for the evolution of similarity coefficients utilises Sneath's SLCA algorithm in the same way as it was originally proposed by McAuley. However, instead of Jaccard's similarity coefficient, a Genetic Programming evolutionary machine is employed for the construction of a population of potential similarity coefficients. The performance of each of the evolved coefficients is evaluated on a set of cell formation test problems which are used as fitness cases (see Figure F in the Appendix).

The aim is to create a new similarity coefficient that will perform at least as good as the man-made Jaccard's coefficient on the solution of cell-formation optimisation

problems. The operation of the GP-SLCA procedure in pseudo-code form is the following:

Procedure Main
initialise population of randomly generated similarity coefficients
run procedure SLCA for each coefficient
loop
> *loop*
> > *select individuals for crossover or mutation*
> > *apply genetic operators and form new coefficients*
> *until a new generation has been formed*
> *run procedure SLCA for each coefficient*
until termination criterion is true

Procedure SLCA
compute similarity matrix for each fitness case
construct dendrogram for each fitness case
loop for each fitness case
> *create machine cells for the highest level of similarity coefficient*
> *assign parts to machine cells*
> *calculate the fitness value of the cell configuration*
> *if solution is the best recorded so far, best=current solution*
until a single cell has been formed
calculate the sum of best solutions and assign the sum as the fitness of the individual

The characteristics of the GP algorithm used in this paper are the following:

Functions: The four basic arithmetic operations (+, -, ×, % (protected division function)) were used for the manipulation of input variables.

Terminals: The basic elements of Jaccard's similarity coefficient were included in the terminal set. In that way it was hoped that the framework would at least be able to evolve the formula of Jaccard's coefficient. In addition, the variable d_{ij} was included in the set. The definition of d_{ij} is the following:

d_{ij} : number of parts processed by neither machine j nor machine i

Optimisation objective: The maximisation of grouping efficacy, Γ, (Kumar and Chandrasekharan, 1990) was used as the quality measure for the evaluation of potential solutions, since it has been employed by a considerable number of researchers in the field. The value of the grouping efficacy is calculated as follows:

$$\Gamma = 1 - \frac{e_v + e_0}{e + e_v} = \frac{e - e_0}{e + e_v}$$

where:

e: total number of non-zero entries in the m/c matrix

e_0: total number of non-zero entries outside the diagonal blocks (exceptional elements)

e_v: total number of zero entries inside the diagonal blocks (voids)

Fitness cases: The choice of cell formation problems that will form the set of fitness cases is not straightforward. This is due to the fact that the cell formation problem is not parameterised, i.e. its difficulty and structure is usually expressed qualitatively rather than quantitatively. In our experiments we collected 27 different test problems from the cell formation literature (Table 1) and used ten different combinations of them as training sets.

No.	Reference	Size
1	Boctor (1991)	16×30
2	”	16×30
3	”	16×30
4	”	16×30
5	”	16×30
6	”	16×30
7	”	16×30
8	”	16×30
9	Boe & Cheng (1991)	20×35
10	Burbidge (1975)	16×43
11	Carrie (1973)	20×35
12	Chan & Milner (1982)	10×15
13	Chandrasekharan & Rajagopalan (1987)	40×100
14	Chandrasekharan & Rajagopalan (1986)	8×20
15	Chandrasekharan & Rajagopalan (1986)	8×20
16	Chandrasekharan & Rajagopalan (1989)	24×40
17	”	24×40
18	”	24×40
19	”	24×40
20	”	24×40
21	”	24×40
22	Kumar et al. (1986)	23×20
23	Kumar & Vannelli (1987)	30×41
24	Seifoddini (1989)	11×22
25	Stanfel (1985)	14×24
26	Stanfel (1985)	30×50
27	Stanfel (1985)	30×50

Table 1: Test problems used in our experimentation

Additional parameters: The values of the parameters that are needed for the run of the GP algorithm are described in Figure 3 (values as suggested in Koza (1992)).

Parameters	Values
Objective:	evolution of a similarity coefficient that maximises grouping efficacy in binary cell formation problems when SLCA is used as the clustering procedure
Terminal set:	*a, b, c, d*
Function set:	+, -, ×, %
Population size:	500
Subtree crossover probability:	.9
Subtree mutation probability:	.1
Selection:	Tournament selection, size 7
Number of generations:	50
Maximum depth for crossover:	17
Initialisation method:	Ramped half and half

Figure 3. Koza tableau for the GP-SLCA methodology

4 Evolution of general-purpose similarity coefficients

4.1 Results

Twenty runs of the GP-SLCA algorithm were conducted for each experimental training set. The cumulative results of the best coefficients evolved are presented in Table 2. For comparison purposes, the performance of Jaccard's similarity coefficient on the same set of problems is also given. Note that the outlined problems in each column were used as fitness cases for the evolution of the corresponding similarity coefficient. The rest of the problems in the same column formed the validation set.

Results from Table 2 indicate that the GP-SLCA framework was able to evolve coefficients that generalised over the entire set of problems. As a measure of the quality of the generalisation, the mean value of grouping efficacy was calculated and compared with the same value achieved by Jaccard's coefficient (Table 3).

The mean value of grouping efficacy produced by all evolved coefficients was similar to the one produced by Jaccard's coefficient. Coefficient SET10 performed particularly well on the entire set of problems, producing an increase 2.6% on average grouping efficacy in comparison to Jaccard's coefficient. Since the difference in performance is relatively small, further research is needed in order to see if SET10 can be distinguished from Jaccard's coefficient. A winner-takes-all comparison of their relative performance on the test problems is presented in Table 4.

Pr.no.	SET1	SET2	SET3	SET4	SET5	SET6	SET7	SET8	SET9	SET10	Jaccard
1	0.5	0.471	0.451	0.471	0.5	0.467	0.438	0.438	0.490	0.467	0.471
2	0.615	0583	0.586	0.586	0.618	0.571	0.588	0.586	0.611	0.601	0.571
3	0.7	0.698	0.698	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
4	0.474	0.459	0.240	0.455	0.489	0.470	0.409	0.231	0.467	0.475	0.474
5	0.727	0.727	0.727	0.727	0.727	0.727	0.727	0.727	0.727	0.727	0.727
6	0.752	0.752	0.752	0.752	0.752	0.752	0.752	0.752	0.752	0.742	0.752
7	0.579	0.579	0.579	0.579	0.570	0.579	0.579	0.238	0.579	0.568	0.579
8	0.773	0.773	0.773	0.773	0.773	0.773	0.748	0.748	0.774	0.774	0.774
9	0.568	0.412	0.554	0.568	0.520	0.562	0.568	0.568	0.568	0.568	0.568
10	0.544	0.556	0.367	0.568	0.545	0.383	0.543	0.552	0.568	0.545	0.544
11	0.760	0.757	0.760	0.760	0.757	0.757	0.760	0.760	0.767	0.757	0.757
12	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92
13	0.840	0.840	0.840	0.840	0.840	0.840	0.840	0.840	0.840	0.84	0.840
14	0.569	0.569	0.569	0.587	0.569	0.587	0.587	0.569	0.587	0.587	0.569
15	0.852	0.852	0.852	0.852	0.639	0.852	0.852	0.852	0.852	0.852	0.852
16	1	1	1	1	1	1	1	1	1	1	1
17	0.851	0.851	0.851	0.851	0.851	0.851	0.851	0.581	0.851	0.851	0.851
18	0.735	0.735	0.735	0.735	0.735	0.735	0.735	0.735	0.735	0.735	0.735
19	0.443	0.532	0.149	0.465	0.503	0.523	0.522	0.136	0.513	0.507	0.517
20	0.454	0.472	0.194	0.466	0.458	0.477	0.309	0.136	0.295	0.453	0.199
21	0.410	0.429	0.330	0.429	0.41	0.431	0.203	0.41	0.203	0.429	0.232
22	0.283	0.246	0.430	0.246	0.246	0.246	0.479	0.246	0.385	0.337	0.246
23	0.520	0.525	0.6	0.528	0.558	0.530	0.6	0.543	0.516	0.585	0.552
24	0.709	0.677	0.731	0.682	0.682	0.650	0.731	0.731	0.720	0.682	0.682
25	0.671	0.7	0.718	0.671	0.699	0.710	0.710	0.706	0.666	0.696	0.671
26	0.558	0.570	0.571	0.567	0.558	0.482	0.468	0.573	0.521	0.561	0.565
27	0.176	0.298	0.484	0.244	0.161	0.224	0.165	0.479	0.147	0.479	0.389

Table 2: Cumulative results on test problems

	SET1	SET2	SET3	SET4	SET5	SET6	SET7	SET8	SET9	SET10	Jaccard
$\overline{\Gamma}$	0.629	0.629	0.610	0.630	0.621	0.622	0.622	0.584	0.621	0.646	0.620

Table 3: Mean value of grouping efficacy

	Jaccard's coefficient better	Jaccard's coefficient worse	Jaccard's coefficient equal
SET10	5	10	12

Table 4: Jaccard's coefficient vs. SET10 in terms of non-dominated solutions

It is obvious that there is a large number of problems where the same level of grouping efficacy has been achieved by both coefficients, thus we cannot safely reject the hypothesis that the two coefficients are the same. The Analysis Of Variance (ANOVA) between the two sets of values confirms this statement (Table 5).

SUMMARY

Groups	Count	Sum	Average	Variance
SET10	27	17.438	0.645852	0.027259
JACCARD	27	16.737	0.619889	0.041745

ANOVA

F	P-value	F crit
0.263753	0.609729	4.026631

Table 5: ANOVA for SET10 and Jaccard's coefficient (α=0.05)

From the *F*-value it is obvious that we cannot reject the null hypothesis. From the above results we can safely conclude that the GP-SLCA algorithm was able to evolve similarity coefficients that performed at least as good as the similarity coefficient devised by human intuition. It will be interesting to take a closer look at the structure of the evolved similarity coefficients. Coefficient SET4 is calculated using the following formula:

$$ d\left(d + c + \frac{b}{c} + a^2\right) - \left(\frac{d}{\frac{b}{c} - a}\right) $$

Notice that GP evolves structures that do not follow the elegant form of Jaccard's coefficient, but are just as effective. From the above formula it is clear that the value of the coefficient is proportional to the values of *a* and *d*. This is expected since these values are indicative of the similarity of parts processed between a pair of machines. The structure of coefficient SET10 is much more complicated, as the following formula depicts:

$$ 4d + 3b - b^2 - ab + ad + \left[\left(\frac{FACTOR - d - b - (d + b)cb + bc + ab}{d(a + b)}\right)\right] - \left(\frac{c - c}{d + b}\right) - \left(\frac{ad + d + b}{b^2c^2 - a - b^2}\right) $$

where FACTOR :

$$ -a + (bc - db - cd - 2a + b)\left[d - \left(\frac{(2a - 3b)(a - b) + ab + db - dc - a + c + \left(\frac{ad}{a}\right)}{a}\right) + \left(\frac{ad}{\left(\frac{a - b}{d - a}\right)/ab^2c}\right)\right] $$

The size and the complexity of this particular coefficient makes the task of explaining its operation quite difficult. It is still evident that the value of the coefficient is proportional to the values of *a* and *d*, however, a number of control terms are also present which seem to fine tune its value in particular fitness cases. Note that there are two terms that according to common algebra should have been simplified:

$$ \frac{c - c}{d + b} \quad \text{and} \quad \left(\frac{ad}{a}\right) $$

However, due to the operation of the protected division function, these expressions will evaluate to '1' if the denominator is equal to '0', which is not an unlikely case. Thus, they should be considered in this form during the calculation of the coefficient value.

The generalisation of this coefficient is quite good. While other coefficients failed to generalise in specific test problems, SET10 seems to have captured information that is relevant to the solution of the problem. In problems 19-21, where the m/c matrices have been custom designed to be difficult for grouping, Jaccard's coefficient failed to find fit partitions. On the same problems SET10 created cell configurations with much higher levels of grouping efficacy. On the notorious problem 27, where evolved coefficients either completely failed to generalise, or their good performance was not mirrored on the set of validation problems, SET10 produced an excellent level of grouping efficacy. While the difference in performance between SET10 and Jaccard's coefficient could not be mathematically confirmed, results on specific test problems indicate that SET10 was able to handle ill-structured matrices in a more efficient way.

5 Conclusions

In this paper we investigated the possibility of evolving new similarity coefficients for the solution of simple cell formation problems, when SLCA is used as the clustering procedure.

The proposed methodology employed a GP algorithm for the creation of potential similarity coefficients. A variety of test problems comprised the training sets that were used for the evolution of the coefficients. The GP-SLCA procedure was able to evolve a number of coefficients that were competitive to a coefficient devised by human intuition (Jaccard's coefficient). While one of the evolved coefficients seemed to generalise better on ill-structured test problems than Jaccard's coefficient, their performance could not be distinguished statistically. However, there is potential for the evolution of better coefficients since the GP system used in this paper can be enhanced in terms of the population size, if more computational power becomes available. In addition, the introduction of parsimony penalties and modularization procedures, like Automatic

Defined Functions (ADF's) (Koza, 1994) can reduce the size of evolved coefficients and increase our understanding of their operation.

Acknowledgments

Bibliography

Boctor, F.F. (1991), "A linear formulation of the machine-part cell formation problem", *Int.J. of Production Research*, vol.29, no.2, pp.343-356.

Boe, W.J., and Cheng, C.H. (1991), "A close neighbour algorithm for designing cellular manufacturing systems", *Int.J. of Production Research*, vol.29, no.10, pp.2097-2116.

Burbidge, J.L. (1975), *The Introduction of Group Technology*, Halste Press, John Wiley, New York, U.S.A.

Carrie, A.S., (1973), "Numerical taxonomy applied to group technology and plant layout", *Int.J. of Production Research*, vol.11, no.4, pp.399-416.

Chan, H.M., and Milner, D.A. (1982), "Direct clustering algorithm for group formation in cellular manufacture", *Journal of Manufacturing Systems*, vol.1, no.1, pp.65-75.

Chandrasekharan, M.P., and Rajagopalan, R. (1986), "MODROC: An extension of rank order clustering for group technology", *Int.J. of Production Research*, vol.24, no.5, pp.1221-1233.

Chandrasekharan, M.P., and Rajagopalan, R. (1987), "ZODIAC - an algorithm for concurrent formation of part families and machine-cells", *Int.J. of Production Research*, vol.25, no.6, pp.835-850.

Chandrasekharan, M.P., and Rajagopalan, R. (1989), "GROUPABILITY: an analysis of the properties of binary data matrices for group technology", *Int.J. of Production Research*, vol.27, no.6, pp.1035-1052.

Dimopoulos, C., and Mort, N. (1999), "A genetic programming methodology for the solution of the cell formation problem", Research Report no.759, Department of Automatic Control & Systems Engineering, University of Sheffield.

Dimopoulos, C., and Zalzala AMS. (1999), "Recent developments in evolutionary computation for manufacturing optimisation: problems, solutions and comparisons", IEEE Transactions in Evolutionary Computation, in print.

Koza, J.R. (1992), *Genetic Programming: On the programming of computers by means of natural selection*, MIT Press, Cambridge.

Koza, J.R, (1994), *Genetic Programming II: Automatic Discovery of reusable programs*, MIT Press, Cambridge, 1994.

Kumar, C.S., and Chandrasekharan, M.P. (1990), "Grouping efficacy: a quantitative criterion for goodness of block diagonal forms of binary matrices in group technology", Int.J. of Production Research, vol.28, no.2, pp.603-612.

Kumar, K.R., and Vannelli, A. (1987), "Strategic subcontracting for efficient disaggregated manufacturing", Int.J. of Production Research, vol.25, no.12, pp.1715-1728.

Kumar, K.R., Kusiak, A., and Vannelli, A. (1986), "Grouping of parts and components in flexible manufacturing systems", *European Journal of Operational research*, vol.24, pp.387-397.

Lee, H., and Garcia-Diaz, A. (1993), "A network flow approach to solve clustering problems" *Int.J. of Production Research*, vol.31, no.3, pp.603-612.

McAuley, J. (1972), "Machine grouping for efficient production", Production Engineer, vol.51, no.2, pp.53-57.

Mitrovanov, S.P. (1966), *The Scientific Principles of Group Technology*, National Lending Library Translation, Boston Spa, Yorkshire, U.K.

Sarker, B.R., and Mondal, S. (1999), "Grouping efficiency measures in cellular manufacturing: a survey and critical review", Int.J. of Production Research, vol.37, no.2, pp.385-314.

Seifoddini, H (1989), "Single linkage vs. average linkage clustering in machine cells formation application", *Computers & Industrial Engineering*, vol.16, pp.419-426.

Selim, M.H., Askin, R.G., and Vakharia, A.J. (1998), "Cell formation in group technology: review, evaluation and directions for future research", Computers & Industrial Engineering, vol.34, no.1, pp.3-20.

Singh, N. (1993), "Cellular manufacturing systems: an invited review", *European Journal of Operational Research*, vol.69, pp.284-291.

Sneath, P.H.A, (1957), "The application of computers to taxonomy", J.of Gen.Microb., vol.17, pp.201-206.

Stanfel, L.E. (1985), "Machine clustering for economic production", *Engineering Costs & Production Economics*, vol.9, pp.73-81.

APPENDIX

McAuleys clustering algorithm

	p1	p2	p3	p4	p5
m1	1	0	1	0	0
m2	0	1	0	1	1
m3	1	0	1	0	0
m4	1	1	0	1	0

Figure A: Example m/c matrix for the illustration of SLCA

$$S_{1,3} = \frac{2}{2+0+0} = 1 \qquad S_{1,4} = \frac{1}{1+1+2} = 0.25 \qquad S_{1,2} = \frac{0}{0+2+3} = 0$$
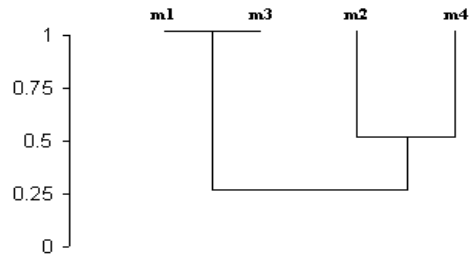
$$S_{2,4} = \frac{2}{2+1+1} = 0.5 \qquad S_{3,4} = \frac{1}{1+1+2} = 0.25 \qquad S_{2,3} = \frac{0}{0+3+2} = 0$$

Figure B: Calculation of similarities using Jaccard's similarity coefficient

	m1	m2	m3
m2	0	*	*
m3	1	0	*
m4	0.25	0.5	0.25

Figure C: Similarity matrix for the example problem

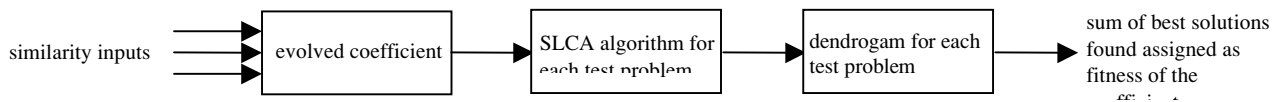**Figure D**: Dendrogram of solutions for the example problem

Solution 1 (initial)   *cell 1*: m_1
                                    cell 2: m_2
                                    cell 3: m_3
                                    cell 4: m_4

Solution 2 (T=1)   *cell 1*: m_1, m_3
                                    cell 2: m_2
                                    cell 3: m_4

Solution 3 (T=0.5)   *cell 1*: m_1, m_3
                                    cell 2: m_2, m_4

Solution 4 (T=0.25)   *cell 1*: m_1, m_2, m_3, m_4

Figure E: Potential solutions

GP-SLCA methodology

Figure F: Illustration of the GP-SLCA procedure for the evolution of similarity coefficients