Massively Parallel Fuzzy Systems: The Case of Three Spiral Pattern Recognition[1]

Sameer Singh, School of Computing, University of Plymouth, UK

Abstract

The main objectives of this paper are: 1) to describe the working of a Massively Parallel Fuzzy System; 2) to test the system on a new benchmark, the three spiral data set; and 3) to describe the behaviour of the system when solving the problem. The described system is aimed at solving pattern recognition problems in real-time. Pattern recognition data are subjected to non-iterative decision making through the estimation of class membership of test data. This paper describes the performance of this system on the temporal three spiral benchmark. The task is to learn three class data which lies on three distinct spirals that coil around each other and around the origin with time. There are no linear solutions to this problem (33% recognition rate with discriminant analysis). The system under consideration classifies the training set with 100% success and recognises training data with up to 89% (without the use of rejection threshold) and 98% success (with a rejection threshold of θ =.8). The behavioural aspects of the system are also studied by quantifying the recognition rate as a function of spiral radius and rejection threshold.

1. Introduction

The two spiral benchmark has been extensively tested to verify several neural architectures and statistical pattern recognition methods. The task is simple: distinguish data of two classes that lies on two distinct spirals that coil around each other and around the origin with time. The spirals are completely symmetrical and it is possible to vary the complexity of the task by varying spiral density φ, radius σ and offset δ. This problem has several interesting features: 1) the spirals coil with time and thus Gaussian methods estimating data density are not very successful; 2) the spiral problem can be made easy or complicated by varying the offset and radius parameter to increase or decrease data density; 3) the spiral problem can not be solved linearly; and 4) previous studies with neural network solutions suggest a range of limitations when using training methods including backpropagation [19].

Unfortunately, previous research on the recognition of temporal data in real-time has been limited. Bishop [1] notes that validation methods are not very well developed for temporal data analysis. In addition, though very important, there is a limited amount of work on data recognition for temporal processes that themselves evolve with time. The spiral data can be found in several applications, e.g. industrial [3], and at present real-time analysis solutions are virtually non-existent. Spiral data classification in two dimension is a difficult task for several pattern recognition algorithms. Several applications involve the same task in more than two dimension. It is therefore important to address the following issues for temporal data: 1) their classification; 2) validation methods; 3) generalisation of techniques across different tasks. Especially for the spiral problem, it would also be required to try and solve the problem in *n* dimension, where $n > 2$.

Currently, linear statistical approaches to classifying two spiral data culminate in decision making by chance (50% success). Neural network solutions are tedious and require several attempts at arriving at an optimal architecture [6,7,10]. Some previous work has attempted to solve the above problem in real-time by either: data encoding [5,8], using novel algorithms, hypercube separation algorithm [20], neurofuzzy systems [17] and novel neural network architectures [4,18]. The solutions are however deficient because: 1) test results are not always encouraging; 2) the benchmark has not been rigorously tested with varying offset and radius parameters; and 3) data not lying immediately on the spiral can not be properly classified [19]. In addition, there is no guarantee that a proposed method will generalise from the two spiral case to a *n* spiral benchmark. In this paper a three spiral benchmark is described in brief which extends the two spiral benchmark. This benchmark will be tested with a new classifier MPFS which has been previously successful in applications including: electronic nose data [14,15] and manufacturing assembly data [12,13]. In

[1] Singh, S. "Massively Parallel Fuzzy Systems: The case of the three spiral recognition", *Proc. IEEE International Conference on Fuzzy Systems FUZZ-IEEE'98, 1998 IEEE World Congress on Computational Intelligence*, Anchorage, Alaska, vol. 2, pp 1066-1072(4-9 May , 1998)

addition, with the two spiral data, previous test results with MPFS approach show success rates of 100% on the validation set and 99.6% success on the test set. The system was also found to be robust to increase in offset which is used to displace the test set: for $.1 \leq \delta \leq 2.5$ with a constant radius $\sigma = 6.5$, recognition rates > 90% in all cases, and for radius changes $.5 \leq \sigma \leq 6.5$ with a constant offset $\delta = .1$, recognition rates > 95% on the test set (training set (**x**, **y**), validation set (**x**, **y**+δ), test set (**x**+δ, **y**)).

2. Three Spiral Benchmark[2]

The three spiral benchmark extends the two spiral problem to three dimension. The third spiral is an average of the first two and coils around the first two and around the origin (Fig.1). The benchmark now has three classes whose data lies on three spirals. This increases the complexity of the two spiral problem and linear solutions are impossible (33% recognition rate with discriminant analysis). Here the recognition rate is the ratio in percentage of the correctly predicted test patterns to the total number of test patterns.

The three spiral benchmark is not only complex to solve, but it also provides an extended set of validation and test sets. For the two spiral benchmark, validation and test sets were generated by offsetting the training data by the amount δ. The same approach is followed for the three spiral benchmark. It is possible to generate a total of eight test sets 3D-*i*; 3D-1 (**x** + δ, **y** + δ, **z** + δ), 3D-2 (**x** + δ, **y** + δ, **z** - δ), 3D-3 (**x** + δ, **y** - δ, **z** + δ), 3D-4 (**x** + δ, **y** - δ, **z** - δ), 3D-5 (**x** - δ, **y** + δ, **z** + δ), 3D-6 (**x** - δ, **y** + δ, **z** - δ), 3D-7 (**x** - δ, **y** - δ, **z** + δ), and 3D-8 (**x** - δ, **y** - δ, **z** - δ) where the training set is {**x**, **y**, **z**}. For neural benchmarking, some of these sets may be used for validation to arrive at an optimal architecture. For statistical classifiers, all sets may be used for test purposes. By varying offset δ and radius σ parameters, a larger number of test sets can be generated. The three spiral training and test sets contain a total of 300 patterns, 100 in each of the three classes.

3. Massively Parallel Fuzzy Systems

Massively Parallel Fuzzy Systems (MPFS) work in a parallel mode. The training data is represented as Ω and the test data as T. A set of processors are arranged in parallel which equal the number of classes under consideration. Processor P_k representing class k calculates the membership of a test pattern in class *k*. Each processor receives the same input pattern at time *t*. The processor which generates the highest membership value claims the class of the test pattern. The membership calculation for a test pattern can be based on the presence of a nearest neighbour in the training set. This nearest neighbour can be identified using *conditional membership* which is defined as: "For a total of n variables, the conditional membership of test datum x_m in class k $\mu_k(x_m)$ where $x_m \in$ T, $1 \leq m \leq n$ depends on all $\mu_k(x_i)$ that are previously known". Here $\mu_k(x_i)$ is the membership of test datum x_i (a part of the test pattern $p_\Gamma = (x_1,.. x_i ... x_n)$) in class k and i varies from 1 to m-1. For a three class problem with X, Y and Z measurements, the possibility of Y and Z is conditional. This implies that for any test pattern $\mu_\Gamma = (X_T, Y_T, Z_T)$ for $1 \leq j \leq 300$, the calculation of $\mu_k(Y_T)$ depends on $\mu_k(X_T)$ and $\mu_k(Z_T)$ depends on the previous calculation of $\mu_k(X_T)$ and $\mu_k(Y_T)$ where k is either class C_1 , C_2 , or C_3. The procedure includes the following steps.

[1] Label the data as for training set Ω, and test set T. The benchmark consists of a total of one training file with 300 patterns and eight test files for offset δ (3D-1 to 3D-8) each containing 300 patterns in each of these data sets, 100 for each class. Start with the first test file.

[2] Separate class C_1, C_2, and C_3 training data in three training files *f1*, *f2*, *f3*.

[3] For the test pattern $p_\Gamma = (X_T, Y_T, Z_T)$ in T, perform the following steps:

[4] Find the upper and lower bounds of X_T for class C_1 from *f1*. These may be represented as X_Ω(lb) and X_Ω (ub). Here lb and ub are positions at which lower and upper bound are found in the training data array. If $X_T > X_\Omega$ for all $X_\Omega \in$ *f1*, then X_T has only a lower bound. Similarly, if $X_T < X_\Omega$ for all $X_\Omega \in$ *f1*, then X_T has only an upper bound.

[5] For *f1*, calculate class memberships for the following cases:

Case

(i) test pattern (X_T, Y_T, Z_T) already exists in the training set

$\mu_1(X_T) = 1.0$; $\mu_1(Y_T) = 1.0$; $\mu_1(Z_T) = 1.0$;

(ii) test pattern does not exist but X_T exists at position j in Ω

$\mu_1(X_T) = 1.0$; $\mu_1(Y_T)) = 1.0/ (1.0 + |Y_T) - Y_j|)$;
$\mu_1(Z_T) = 1.0/ (1.0 + |Z_T) - Z_j|)$;

(iii) *test pattern does not exist but Y_T exists at position j in Ω*

[2] The spiral benchmark is an extension of the two spiral benchmark available at the Carnegie Mellon AI repository.

$\mu_1(Y_T) = 1.0$; $\mu_1(X_T) = 1.0/ (1.0 + |X_T - X_j|)$;
$\mu_1(Z_T) = 1.0/ (1.0 + |Z_T - Z_j|)$;

(iv) *test pattern does not exist but Z_T exists at position j in Ω*

$\mu_1(Z_T) = 1.0$; $\mu_1(X_T) = 1.0/ (1.0 + |X_T - X_j|)$;
$\mu_1(Y_T) = 1.0/ (1.0 + |Y_T - Y_j|)$;

(v) None of the test values in the test pattern (X_T, Y_T, Z_T) exist in the training set Ω

$\eta_1 = |X_\Omega(ub) - X_T|$
$\eta_2 = |X_T - X_\Omega(lb)|$
$\eta_3 = |Y_\Omega(ub) - Y_T|$
$\eta_4 = |Y_T - Y_\Omega(lb)|$
$\eta_5 = |Z_\Omega(ub) - Z_T|$
$\eta_6 = |Z_T - Z_\Omega(lb)|$

if($\eta_1*\eta_3*\eta_5 < \eta_2*\eta_4*\eta_6$)
{
$\mu_1(X_T) = 1.0/ (1.0 + \vartheta(\eta_1))$; ... (1)
where ϑ is a function to weigh the role of η. We have used the square root function in this study.
$\mu_1(Y_T) = 1.0/ (1.0 + \vartheta(\eta_3 + \vartheta(\eta_1)))$; ... (2)
$\mu_1(Z_T) = 1.0/ (1.0 + \vartheta(\eta_5 + \vartheta(\eta_3 + \vartheta(\eta_1))))$; ... (3)
}

if($\eta_2*\eta_4*\eta_6 < \eta_1*\eta_3*\eta_5$)
{
$\mu_1(X_T) = 1.0/ (1.0 + \vartheta(\eta_2))$;
$\mu_1(Y_T) = 1.0/ (1.0 + \vartheta(\eta_4 + \vartheta(\eta_2)))$;
$\mu_1(Z_T) = 1.0/ (1.0 + \vartheta(\eta_6 + \vartheta(\eta_4 + \vartheta(\eta_2))))$;
}

When X_T has only an upper bound, or only a lower bound, the above calculations are adjusted for this to exclude the missing information.

[6] The membership of a test pattern in class C_1, $\mu_1(X_T, Y_T, Z_T)$ is given by:
$\mu_1(X_T, Y_T, Z_T) = \xi(\mu_1(X_T), \mu_1(Y_T), \mu_1(Z_T))$ where ξ function is used to combine membership values. The two most commonly used functions used are the min-max function and the product operator [2]. In this study we have used the multiplication operator.

[7] Perform the steps 4 to 6 on *f1* and *f2* to calculate $\mu_2(X_T, Y_T, Z_T)$ and $\mu_3(X_T, Y_T, Z_T)$.

[8] If $\mu_1(X_T, Y_T, Z_T) > \mu_2(X_T, Y_T, Z_T)$ and $\mu_1(X_T, Y_T, Z_T) > \mu_3(X_T, Y_T, Z_T)$ then test pattern $(X_T, Y_T, Z_T) \in C_1$
If $\mu_2(X_T, Y_T, Z_T) > \mu_1(X_T, Y_T, Z_T)$ and $\mu_2(X_T, Y_T, Z_T) > \mu_3(X_T, Y_T, Z_T)$ then test pattern $(X_T, Y_T, Z_T) \in C_2$
If $\mu_3(X_T, Y_T, Z_T) > \mu_1(X_T, Y_T, Z_T)$ and $\mu_3(X_T, Y_T, Z_T) > \mu_2(X_T, Y_T, Z_T)$ then test pattern $(X_T, Y_T, Z_T) \in C_3$

[9] Go to step 3 for the next test pattern analysis.

[10] Calculate the total number of test patterns correctly predicted belonging to their class. The recognition rate *R* is the proportion to the total number of correctly predicted patterns to the total size of the test space, i.e. the number of patterns in T.

The conditional memberships $\mu_1(Y_T)$ and $\mu_1(Z_T)$ depend on the upper and lower bounds identified for calculating $\mu_1(X_T)$. The three memberships may be differentiated with respected to the distance η_1 which allows us to describe the classifier behaviour. If η_1 is large then $\mu_1(X_T)$ is small and it has a considerable effect on later variables Y and Z, though this effect diminishing as η_1 is nested. We may derive the change in all three memberships with respect to η_1.

$$\frac{dy}{dx} = \frac{-1}{2\sqrt{x} + 2x\sqrt{x} + 4x} \quad \text{... (4)}$$

where y is $\mu_1(X_T)$ given by eqn. (1) and *x* is η_1.

$$\frac{dy}{dx} = \frac{-1}{4(1+\sqrt{\eta_3 + \sqrt{x}})^2 . \sqrt{\eta_3 + \sqrt{x}} . \sqrt{x}} \quad \text{... (5)}$$

where y is $\mu_1(Y_T)$ given by eqn. (2) and *x* is η_1.

$$\frac{dy}{dx} = \frac{-1}{8(1+\sqrt{\eta_5 + \sqrt{\eta_3 + \sqrt{x}}})^2 . (\sqrt{\eta_5 + \sqrt{\eta_3 + \sqrt{x}}}) . (\sqrt{\eta_3 + \sqrt{x}}) . \sqrt{x}} \quad \text{...(6)}$$

where y is $\mu_1(Z_T)$ given by eqn. (3) and *x* is η_1.

It is possible to plot these rate of change function values to explain MPFS behaviour. The behaviour may be studied by calculating changes in conditional class memberships as a function of distance η_1, i.e. plotting dy/dx for test patterns. This is shown in Fig. 2 for data generated with threshold $\theta = .8$, radius $\sigma = 96.5$ and offset $\delta = .1$. The natural logarithm of dy/dx is plotted on the Y axis for the first 51 patterns in the test set. A higher degree of change occurs in the top half of the spectrum since negative values have been plotted. It may be seen that the change itself is similar to the spiral structure. As expected, the rate of change in $\mu_1(X_T)$ is most dependent on η_1 compared to

$\mu_1(Z_T)$. Also the rate of change in $\mu_1(X_T)$ is most pronounced for patterns 1, 4, ... which should be classed as of type C_1. The same is true for $\mu_1(Y_T)$ for patterns 2, 5, ... and for $\mu_1(Z_T)$ for patterns 3, 6, ... Since the test data is interleaved, class k patterns are in positions k, k+3, ...; this further proves that the MPFS system is responsive to the three spiral benchmark. It produces possibilities whose rate of change is data dependent and therefore results in a high recognition rate.

4. Results

Table 1 shows the recognition rate in percentage as a function of the radius σ of the spiral. The recognition rate increases rapidly at first, and then saturates for higher σ. By increasing the radius at first, spiral data is spread over a larger input space thus increasing the probability of its correct classification. However, after a certain radius which defines the envelope of the spiral generated, σ =36.5, recognition rate increases very slowly. This behaviour is generic across different test sets 3D-*i*. The recognition rate achieved finally, R = 89%, is considerably good and very encouraging for further studies with MPFS. The MPFS behaviour can also be studied by understanding how decisions are made. Bishop [1] notes that: "In general we expect most of the misclassification errors to occur in those regions of **x**-space where the largest of the posterior probabilities is relatively low, since there is a strong overlap between different classes," (p. 28). The same approach can be used for possibilistic decision making in the MPFS system. We may only consider those patterns for decision making where at least one of the possibility is greater than a pre-defined threshold θ. Hence, for the three spiral benchmark, the decision space will consist of only those test patterns T for which at least one of the memberships, $\mu_k(X_T)$, $\mu_k(Y_T)$, $\mu_k(Z_T)$ is greater than the rejection threshold θ. In Table 2 the decision space (ratio of the patterns meeting the criteria to the total number in the test set, i.e. 300, in percentage) is shown in italics. The recognition rate is shown as a percentage of correctly classified pattern with respect to the total number of qualifying patterns. Table 2 shows that as the threshold θ is increased, the decision space shrinks but as expected, [1], the recognition rate improves at the same time as there is little possibility overlap between classes. It may be useful for MPFS to only make decisions for $\theta > .5$ or some other pre-defined limit in order to improve the quality of decision making. It may be seen that the three spiral problem can be recognised with 89% success if all patterns are included for test, and with 98% success if only 81% of the total number test patterns are included, i.e. 243 patterns. The experimental results were produced in real-time. On a Pentium 200 Mhz machine, the train-test cycle for testing 300 patterns took less than 2 minutes (each patterns was tested in less than 0.5 seconds).

5. Conclusion

The above results are very encouraging. The three spiral benchmark presents a new challenge for testing novel pattern recognition architectures on temporal data. The results show that the MPFS system solves the task in real-time with nearly 89% recognition success on all test sets and nearly 98% success with the use of rejection thresholds. Currently, the three spiral benchmark has not been used with neural networks so a comparative account of MPFS performance is not possible. MPFS systems however seem robust and reliable classifier systems as indicated by Figures 1 & 2. Domain specific membership calculation methods will be needed for further studies on the spiral benchmark since Gaussian functions, or those replicating them in other forms [11,22] are not suitable for temporal decision making. It is also necessary to generate appropriate validation procedures for spiral data [1,16,21].

References

[1] C. M. Bishop, *Neural networks for pattern recognition*, Clarendon Press, Oxford, 1995.

[2] M. Brown and C. Harris, *Neurofuzzy adaptive modelling and control*, Prentice Hall, 1994.

[3] M. Brown and C. A. Harris, "Fuzzy logic expert system for iron ore processing," *Proceedings of the IEEE industry applications conference*, Canada, vol. 3, pp. 2190-9, 1993.

[4] T. Chin-Chi and B. W. Wah, "An automated design system for finding the minimal configuration of a feed-forward neural network," *Proceedings of the IEEE International conference on neural networks*, Florida, vol. 3, pp. 1295-1300, 1994.

[5] H. C. Chua, J. Jia, L. Chen and Y. Gong, "Solving the two-spiral problem through input data encoding," *Electronics letters*, vol. 31, issue 10, 813-14, 1995.

[6] S. E. Fahlman, "Faster-learning variations on back-propagation: An empirical study," *Proceedings of the 1988 Connectionist models summer school*, Morgan Kaufmann, 1988.

[7] S. E. Fahlman and C. Lebiere, "The Cascade-Correlation learning architecture," In Touretzky (Ed.) *Advances in neural information processing systems 2*, Morgan Kaufmann, 1990.

[8] J. Jia and H. Chua, "Solving two-spiral problem through input data representation," *Proceedings of the IEEE International conference on neural networks*, vol 1, pp. 132-135, 1995.

[9] B. Kosko, *Neural networks and fuzzy systems - A dynamical systems approach to Machine Intelligence*, Prentice Hall International edition, 1992.

[10] K. J. Lang and M. J. Witbrock, "Learning to tell two spirals apart," *Proceedings of the 1988 Connectionist models summer school*, Morgan Kaufmann, 1988.

[11] *Fuzzy reasoning and its applications*, E. H. Mamdani and B. R. Gaines (eds.), Academic Press, 1981.

[12] S. Singh and M. Steinl, "Fuzzy search techniques in knowledge-based systems," *Proceedings of the 6th International conference on data and knowledge systems for manufacturing and engineering (DKSME'96)*, Tempe. Arizona, pp. 1-10, 1996.

[13] S. Singh and M. Steinl, "Fuzzy pattern recognition for knowledge-based systems," *Proceedings of the International conference on knowledge based computer systems*, Bombay, 1996, pp. 383-394.

[14] S. Singh, E. L. Hines and J. Gardner, "Fuzzy neural computing of coffee and tainted water data on electronic nose," *Sensors and Actuators B*, vol. 30, issue 3, pp. 190-195, 1996.

[15] S. Singh, E. L. Hines and J. Gardner, "Classifier systems based on possibility distributions: A comparative study," *Proceedings of the 3rd International conference on neural networks and genetic algorithms (ICANNGA97)*, Norwich, Wien: Springer-Verlag, 1997.

[16] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *Journal of the Royal Statistical Society*, 36, pp. 111-147, 1974.

[17] C. T. Sun and J. S. Jang, "A neuro-fuzzy classifier and its applications," *Proceedings of the IEEE International conference on fuzzy systems*, vol. 1, pp. 94-98, 1993.

[18] L. P. Tay and D. J. Evans, "Fast learning artificial neural network (FLANN II) using the nearest neighbour recall," *Neural, parallel and scientific computations*, vol. 2, issue 1, pp.17-27, 1994.

[19] D. S. Touretzky and D. A. Pomerleau, "What's hidden in the hidden layers ?" *Byte* (August issue), pp. 227-233, 1989.

[20] F. Ulgen, N. Akamatsu and T. Iwasa, "The hypercube separation algorithm: a fast and efficient algorithm for on-line handwritten character recognition," *Applied Intelligence*, vol. 6, issue 2, 101-116, 1996.

[21] S. M. Weiss, and C. A. Kulikowski, *Computer systems that learn*, Morgan Kaufmann, San Mateo, CA, 1991.

[22] L. A. Zadeh, *Fuzzy logic and its applications*, Academic press, New York, 1965.

Table 1. 3D spiral recognition rate % as a function of the radius σ with an offset δ = 0.1. The recognition rate is represented as a function of the total number of correctly predicted patterns.

radius σ	*3D-1*	*3D-2*	*3D-3*	*3D-4*	*3D-5*	*3D-6*	*3D-7*	*3D-8*
2.5	38	36	35	35	35	34	34	33
**6.5*	51	50	50	51	55	54	54	54
16.5	69	70	68	69	70	70	70	70
26.5	79	78	78	79	79	78	78	78
36.5	85	84	84	84	84	84	84	83
46.5	85	85	85	85	85	85	85	84
56.5	88	87	87	87	87	87	87	87
66.5	89	88	88	88	87	88	87	87
76.5	89	89	88	88	87	88	87	87
86.5	89	89	88	88	87	88	87	87
96.5	89	89	89	88	87	88	87	88

Table 2. 3D spiral recognition rate % and decision space % as a function of rejection threshold θ for radius σ = 96.5 and offset δ = 0.1.

rejection threshold θ	*3D-1*	*3D-2*	*3D-3*	*3D-4*	*3D-5*	*3D-6*	*3D-7*	*3D-8*
0	89 *100*	89 *100*	89 *100*	88 *100*	87 *100*	88 *100*	87 *100*	88 *100*
.2	94 *87*	94 *86*	94 *87*	94 *87*	95 *86*	94 *87*	95 *86*	94 *87*
.4	96 *85*	95 *85*	95 *85*	95 *85*	96 *85*	96 *85*	96 *85*	96 *85*
.6	97 *84*	97 *84*	97 *84*	98 *84*	98 *84*	98 *84*	98 *84*	97 *84*
.8	97 *82*	97 *82*	97 *82*	97 *82*	98 *81*	98 *81*	98 *81*	98 *81*

Figure 1. The 3 spiral data in two dimensions: series1 ≡ X dimension; series2 ≡ Y dimension; series3 ≡ Z dimension.

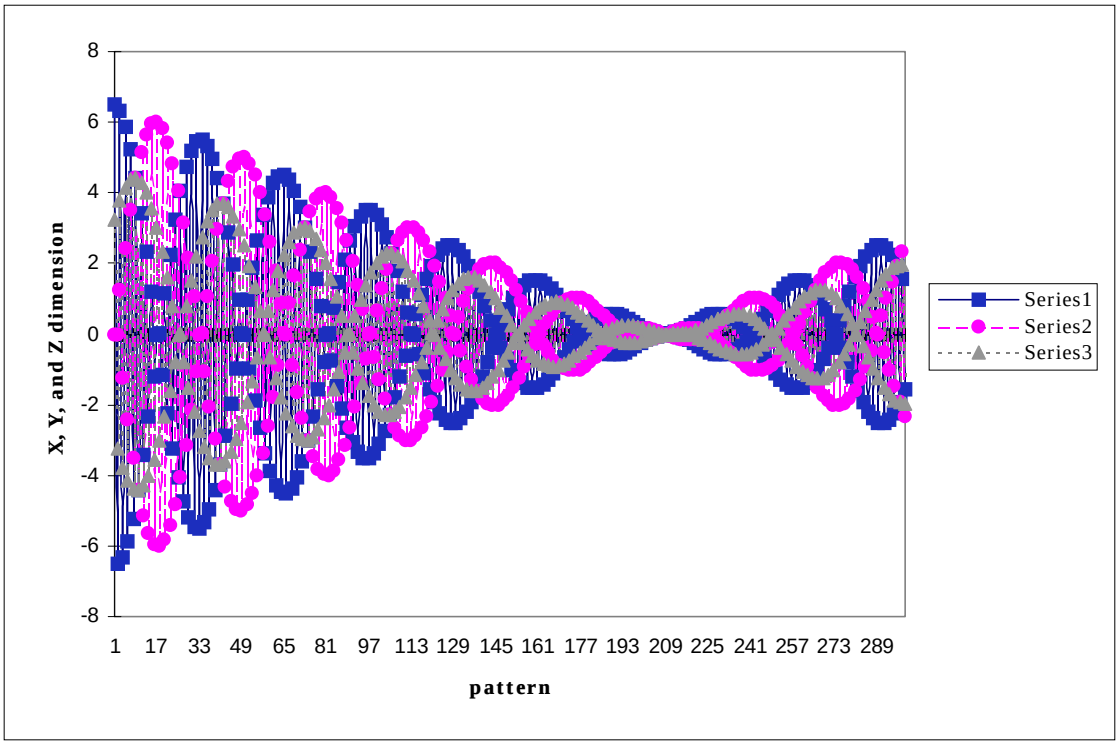

Figure 2. Rate of change of possibilities $\mu_1(X_T)$ ≡ series1, $\mu_1(Y_T)$ ≡ series2, and $\mu_1(Z_T)$ ≡ series3, for class C_1 with the first 51 patterns. See eqns. (4), (5) & (6).