1 vs. 1

Neural Computation in Stylometry I: An Application to the Works of Shakespeare and Fletcher

ROBERT A. J. MATTHEWS
Oxford, UK

THOMAS V. N. MERRIAM
Basingstoke, UK

Abstract

We consider the stylometric uses of a pattern recognition technique inspired by neurological research known as neural computation. This involves the training of so-called neural networks to classify data even in the presence of noise and non-linear interactions within data sets. We provide an introduction to this technique, and show how to tailor it to the needs of stylometry. Specifically, we show how to construct so-called multi-layer perceptron neural networks to investigate questions surrounding purported works of Shakespeare and Fletcher. *The Double Falsehood* and *The London Prodigal* are found to have strongly Fletcherian characteristics, *Henry VIII* strongly Shakespearian characteristics, and *The Two Noble Kinsmen* characteristics suggestive of collaboration.

1. Introduction

Stylometry attempts to capture quantitatively the essence of an individual's use of language. To do this, researchers have proposed a wide variety of linguistic parameters (e.g. rare word frequencies or ratios of common word usage) which are claimed to enable differences between individual writing styles to be quantitatively determined.

Critics of stylometry rightly point out that despite its mathematical approach the technique can never give incontrovertible results. However, there can be little doubt that the case in favour of attributing a particular work to a specific author is strengthened if a wide variety of independent stylometric tests point to a similar conclusion. The development of a new stylometric technique is thus always of importance, in that it can add to the weight of evidence in support of a specific hypothesis.

To be a useful addition to stylometry, a new technique should be theoretically well-founded, of measurable reliability, and of wide applicability.

In this paper, we introduce a technique that meets all these criteria. Based on ideas drawn from studies of the brain, this so-called neural computation approach forms a bridge between the method by which literary scholars reach their qualitative judgements, and the quantitative techniques used by stylometrists.

Like a human scholar, the technique uses exposure to many examples of a problem to acquire expertise in solving it. Unlike a human scholar, however, the neural computation technique gives repeatable results of measurable reliability. Furthermore, the technique is theoretically well-founded. It can be shown that neural networks are capable of approximating any practically useful function to arbitrary accuracy (see, for example, Hecht-Nielsen, 1990, p. 131). Furthermore, ways of finding such networks have their origins in well-established concepts drawn from the theory of statistical pattern recognition and non-linear regression; indeed, neural computation can be thought of in more prosaic terms as a non-linear regression technique.

In addition, neural networks are known to cope well with both noisy data and non-linear correlations between data, confounding effects that have long dogged stylometric research.

With such attributes, neural computation would seem to constitute a promising new stylometric method. In this paper, we show how to construct a stylometric neural network, and then apply it to the investigation of the works of Shakespeare and his contemporary John Fletcher (1579–1625).

2. Background to Neural Computation

Despite the substantial computational power now available to conventional computers, the brain of an infant can still outperform even the fastest supercomputers at certain tasks. A prime example is that of recognizing a face in a crowd: conventional computing techniques have proved disappointing in such tasks.

This has led to interest in so-called neural computing, which is an attempt to imitate computationally the essentials of neurological activity in the brain. The idea is that problems such as pattern recognition may be better solved by mimicking a system known to be good at such tasks.

Neural computation typically (but not necessarily) involves programming a conventional computer to behave as if it consisted of arrangements of simple interconnected processing units—'neurons'—each one of which is linked to its neighbours by couplings of various strengths, known as 'connection weights'. It is now known that even a relatively crude representation of the collective behaviour of real neurons enables a number of difficult computational problems to be tackled.

To do this, the network of neurons has to be trained to respond to a stimulus in the appropriate way. This requires the application of a 'learning algorithm' enabling the weights to converge to give a network producing acceptable solutions. Thereafter, each time the network receives a specific input, it will produce an output consistent with the data on which it has been trained.

Research into such 'neural computation' began in the

Correspondence: Robert Matthews, 50 Norreys Road, Cumnor, Oxford OX2 9PT, UK.

1940s, but it was not until the mid 1980s and the publication of *Parallel Distributed Processing* (Rumelhart and McClelland, 1986) that the current interest in the field was kindled. This followed the authors' demonstration that a type of learning algorithm known as back propagation (or simply 'backprop') enabled neural networks to solve highly non-linear problems that had defeated simple networks (Minsky and Papert 1969).

The backprop algorithm, which had in fact been previously discovered by several researchers, has since been used to produce neural networks capable of solving an astonishing variety of prediction and classification problems, from credit risk assessment to speech recognition, many of which have proved all but intractable by conventional computational techniques (see, for example, Anderson and Rosenfeld, 1989; Refenes *et al.*, 1993).

The backprop algorithm is typically used in conjunction with a specific arrangement of neurons known as the multilayer perceptron (MLP; see Fig. 1). This consists of an input layer of neurons, a so-called hidden layer, and an output layer. Multi-layer perceptions are currently the most widely used form of neural network. They have proved capable of performing classification and prediction even in the presence of considerable non-linearity and noise in the raw data. It is for these reasons that we decided to investigate the specific use of MLPs as a new stylometric discrimination technique.

3. Building a Stylometric MLP

For our purposes, we require an MLP that can take a set of m stylometric discriminators for a given sample of the works of one of two authors, X and Y, and then classify the input as the work of either X or Y. This implies that the MLP will consist of an input layer of m neurons—one for each stylometric discriminator used to differentiate between the two authors—a hidden layer of n neurons, and an output layer of two neurons, corresponding to the two authors.

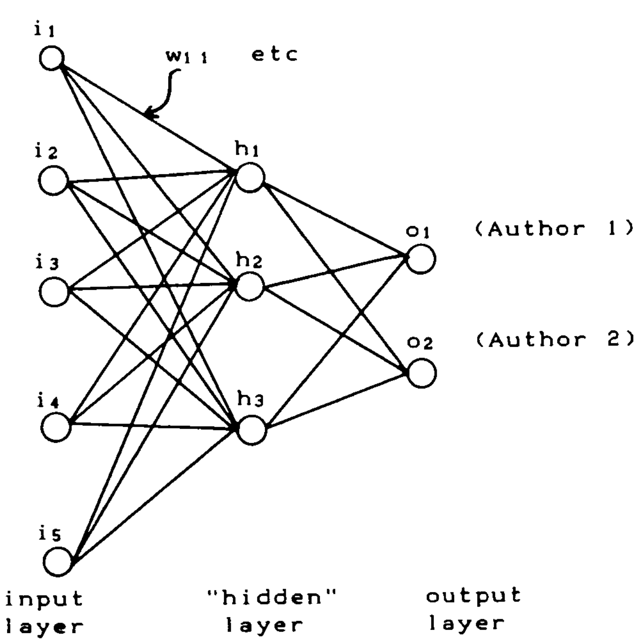

Fig. 1 Topology of a stylometric multi-layer perceptron for classifying works of two authors using five discriminators.

Training such an MLP requires the backprop algorithm, whose derivation is given in Chapter 7 of *Parallel Distributed Processing* (Rumelhart and McClelland, 1986). We then use the following protocol to train the MLP:

(*a*) Prepare k training vectors. These consist of m real numbers representing the discriminators, while the output data consists of the author ID.
(*b*) Set up the weights of the neural network with small random values.
(*c*) Calculate the output that results when the input training vector is applied to this initial network arrangement.
(*d*) Calculate the (vector) difference between what the network actually produces, and the desired result; this constitutes an error vector for this input and output vector.
(*e*) Adjust the weights and thresholds of the network using the backprop algorithm to reduce the error.
(*f*) Repeat with the next input training vector, and continue down the training set until the network becomes acceptably reliable.

We now consider the practical aspects of this protocol.

3.1 *The Training Vectors*

These consist of the m discriminators with the power to differentiate between author X and author Y, together with an author ID label.

In general, the larger m becomes, the stronger the discrimination. However, a limit on the number of discriminators that can be used is set by the the availability of text of reliable provenance on which training can be based. If an MLP has too many inputs relative to the number of training vectors, it will lose its ability to generalize to new data; essentially, there are too many unknowns for the data to support. To combat this, experience shows (D. Bounds, 1993, private communication) that the total number of training vectors used, k, should be at least ten times the sum of the number of inputs and outputs. These training vectors should, moreover, be drawn equally from the works of the two authors, be suitably representative, and be derived from reasonable lengths of text.

The use of many discriminators thus raises the number of training vectors required. However, one can only extract more training vectors from a given amount of reliable training text by taking smaller and smaller samples, and these will be increasingly subject to statistical noise.

Given these various constraints, we concluded that a useful stylometric MLP should consist of five input neurons, giving reasonable discriminatory power, and two outputs; this then leads to a requirement for at least $10 \times (5 + 2) = 70$ training vectors, roughly half of which come from each of the two authors. This number of training vectors allows the stylometric discriminator data to be based on reasonable samples of 1000 words drawn from the core canons of many authors.

3.2 *Training the MLP*

The first step in the training process is the so-called forward pass, in which an input vector is applied to the input neurons, and their output is passed via a set of initially random weights to neurons in the hidden layer. Suppose the discriminators applied to the input layer form the vector (i_1, i_2, i_3, i_4, i_5). Then for each hidden layer neuron h_j we form the sum

$$S(h_j) = \sum_{m=0}^{5} (i_m w_{mj}) \qquad (1)$$

where w_{mj} is the weight connecting input neuron m to hidden layer neuron j. The summation runs from 0 to 5, with w_{0j} the so-called biassing weight which performs a role similar to that of a threshold (Rumelhart and McClelland, 1986, p. 329). It can be trained just like the other weights, with i_0 simply being considered to have the fixed value + 1.

The output from h_j is then obtained by applying a so-called squashing function to S, typically sigmoidal in form, so that

$$\Omega(h_j) = 1/\{1 + \exp[-S(h_j)]\} \qquad (2)$$

These are then used as the inputs to the output layer, with a similar summing and squashing procedure giving $S(o_1)$ and $S(o_2)$ for the two output neurons. The corresponding outputs $\Omega(o_1)$ and $\Omega(o_2)$ constitute the final output of the MLP. Classification is then achieved on the basis of which of these two outputs is the larger.

The error vector, $\boldsymbol{\varepsilon}$, between the desired output and that produced by the network during training is used to modify the weights according to the backprop algorithm. The training is repeated down the training set until the initially random weights converge to the set of values giving an acceptable accuracy of classification. Thereafter the MLP simply uses (1) and (2) to calculate output vectors from given input vectors using the weights w_{mj}, etc., at their converged values.

3.3 *The Completion of Training*

During training, the classification error falls until it reaches a stable value. In practice two criteria are used to dictate when an MLP can be considered 'trained'. Typically, the set of k input vectors is split into a *training* set and a *cross-validation* set. The former is used to train the network while the latter is held in reserve to gauge performance.

Left to train over many cycles, MLPs often learn to classify the training set with complete accuracy. However, this does not imply that the MLP will perform well when exposed to data it has never seen before. This inability to generalize to new data is known as 'overtraining'.

The exact cause of overtraining is still unclear (see, for example, Hecht-Nielsen, 1990 p. 116), but it has obvious symptoms: as training continues, classification of the training vectors continues to improve, while that of the cross-validation vectors start to degrade.

The solution is to halt training when the MLP performs to an acceptable standard on both training and cross-validation vectors. Selecting an appropriate standard is thus a balance between the need to produce useful results and the avoidance of overtraining. Obviously, a 50% success rate in classifying data between two equally likely alternatives is no better than coin-tossing. However, achieving 100% accuracy in both training and cross-validation is usually prevented by the overtraining phenomenon.

We now describe our solution of this and other practical issues surrounding the construction of a stylometric MLP capable of discriminating between Shakespeare and Fletcher.

4. Construction of the Shakespeare–Fletcher MLP

4.1 *Choice of Discriminants*

The inputs of the MLP are the m discriminators we choose as being capable of differentiating between the works of Shakespeare and those of Fletcher. The discriminators should, in addition, show reasonable stability across the corpus of an author's work (at least that made up by works of one genre, such as plays), and ideally maintain their reliability when works are broken down into smaller units, such as individual acts. This latter feature is particularly desirable in an MLP designed to investigate supposed collaborations within a single work.

Both Merriam (1992) and Horton (1987) have studied the choice of discriminators meeting such criteria in considerable detail, and we investigated the use of five discriminants based on their work as inputs for two Shakespeare–Fletcher neural networks.

The Merriam-based set of m = 5 discriminators were the following ratios: *did/(did+do); no/T-10; no/(no+not); to the/to; upon/(on+upon)*. Here *T*-10 is Taylor's ten function words (*but, by, for, no, not, so, that, the, to, with*) (Taylor, 1987).

The set of five discriminators based on the work of Horton consists of ratios formed by dividing the total numbers of words in a sample by the number of occurrences of the following five function words: *are; in; no; of; the*. All contractions involving these function words (e.g. *i' th'*) have been expanded to maximize the word counts.

4.2 *Formation of Training and Cross-validation Data Sets*

For each set of five discriminators, we formed training sets of k = 100 vectors (fifty each for Shakespeare and Fletcher), with each vector taking the following form:

(ratio 1; ratio 2; ratio 3; ratio 4; ratio 5; author ID)

For training purposes, each ratio was computed by word counts on 1,000-word samples from works of undisputed origin for each author. For Shakespeare these were taken to be the core canon plays *The Winter's Tale, Richard III, Love's Labour's Lost, A Midsummer Night's Dream, 1 Henry IV, Henry V, Julius Caesar, As You Like It, Twelfth Night* and *Antony and Cleopatra*. For Fletcher, we took as core canon *The Chances, The Womans Prize, Bonduca, The Island Princess, The Loyal Subject* and *Demetrius and*

Enanthe. For all these, the source used for our word counts was the machine-readable texts produced by the Oxford University Computing Service.

Once the five sets of 100 ratios were extracted for each discriminator, each set was normalized to give zero mean and unit standard deviation to ensure that each discriminator contributes equally in the training process.

4.3 *Training Criteria*

The training vectors thus derived were then used to produce two MLPs: one capable of differentiating between Shakespeare and Fletcher using the five Merriam discriminators, the other using those of Horton.

After some experimentation, it emerged that we could reasonably expect cross-validation accuracies of at least 90% without running into overtraining problems. Thus the first of our criteria for the completion of training was that the MLP be capable of classifying the cross-validation vectors with an accuracy of at least 90%.

The other criterion was set by the requirement that the MLP be unbiassed in its discrimination process; in other words, that it was no more likely to misclassify works of Fletcher as Shakespearian than it was to do the reverse. Thus, the second of our training criteria was that misclassified vectors be approximately equally divided between the two classifications.

These criteria were then used to find a suitable size for the hidden layer. Too few hidden units fails to capture all the features in the data, while too many leads to a failure to generalize; in tests, we found that three hidden units were sufficient to give cross-validation results meeting our criteria. We then fixed our topology for the stylometric MLP at five inputs, three hidden units, and two outputs.

Both the Merriam and Horton MLPs were found to successfully meet the training criteria after twenty or so presentations of the complete 100-vector training set. The Merriam-based network (henceforth MNN) achieved a cross validation accuracy of 90%, with the 10% misclassified being split into 6% Shakespeare classified as Fletcher, and 4% Fletcher classified as Shakespeare.

The Horton-based network (henceforth HNN) achieved 96% cross-validation accuracy, with the both modes of misclassification lying at 2%.

4.4 *Testing and Performance Appraisal*

Having been trained, both MNN and HNN were tested by being asked to classify core canon works of Shakespeare and Fletcher that neither network had seen during training. This constitutes a test of the power of each network to generalize to new data.

In the first test, each network was asked to classify ten complete plays, eight from the core canon of Shakespeare (*All's Well that Ends Well, Comedy of Errors, Coriolanus, King John, Much Ado about Nothing, The Merchant of Venice, Richard II*, and *Romeo and Juliet*) and two from that of Fletcher (*Valentinian* and *Monsieur Thomas*).

In addition to giving the simple (bipolar) classification of 'Shakespeare' or 'Fletcher', as dictated by the larger of the two output signal strengths, each network also provided a measure of the degree to which it considered each work to belong to one class or another. We call this the Shakespearian Characteristics Measure (SCM); it is defined as

$$\mathrm{SCM} = \Omega_S/(\Omega_S + \Omega_F) \quad (3)$$

where Ω_S and Ω_F are the values of the outputs from the Shakespeare and Fletcher neurons, respectively. Thus the stronger the Shakespeare neuron output relative to the Fletcher neuron output, the higher the SCM. Strongly Fletcherian classifications, on the other hand, give SCM closer to zero, and those on the borderline ($\Omega_S = \Omega_F$) give SCM = 0.5. The value of the SCM lies in the greater insight it provides into a particular classification result.

The results obtained from the Merriam and Horton MLPs applied to entire core canon plays of both dramatists are shown in Table 1.

Table 1 Multi-layer perception results for core canon Shakespeare and Fletcher

Play	Merriam SCM	Merriam Verdict	Horton SCM	Horton Verdict
Shakespeare				
ADO	0.75	Shakespeare	0.71	Shakespeare
AWW	0.74	Shakespeare	0.92	Shakespeare
CE	0.90	Shakespeare	0.91	Shakespeare
COR	0.84	Shakespeare	0.98	Shakespeare
KJ	0.76	Shakespeare	0.91	Shakespeare
MV	0.67	Shakespeare	0.97	Shakespeare
R II	0.81	Shakespeare	0.92	Shakespeare
ROM	0.80	Shakespeare	0.87	Shakespeare
Fletcher				
VAL	0.46	Fletcher	0.30	Fletcher
MTH	0.32	Fletcher	0.29	Fletcher

As can be seen, both MNN and HNN gave the correct overall classification to all ten complete plays. The two networks also gave SCMs of similar numerical value, despite being based on different sets of discriminator: the correlation coefficient between the SCMs produced by the two MLPs is 0.894.

The statistical significance of the overall classification results can be judged by using the binomial distribution to calculate the probability $P(S)$ of obtaining at least S successes in T trials simply by chance, given two equally likely outcomes. In our case, we have $T = 10$ and $S = 10$, so that $P(10) = 9.8 \times 10^{-4}$; the correct classification of ten entire plays by both MNN and HNN is thus highly significant ($P < 0.001$).

The significance of the correlation of SCMs can be assessed using the Student t-test, which for $r = 0.894$ and eight degrees of freedom gives $t = 5.653$, corresponding to $P < 0.001$.

These impressive results highlight an important feature of stylometric MLPs: although each network was trained to give 90% cross-validation accuracy, this figure can be improved upon when the networks are applied to entire plays. This reflects the fact that discriminator values derived from entire plays are less noisy than those derived from acts.

We would, however, expect the performance of the MLPs to be somewhat less impressive when they are applied to individual acts, whose stylometric properties will be rather more noisy.To investigate this degradation in performance, we used MNN and HNN to classify individual acts of two plays from the core canon of each playwright. For Shakespeare, we took the acts from *The Tempest* and *The Merry Wives of Windsor*, while for Fletcher we took acts from *Valentinian* and *Monsieur Thomas*.

The Merriam-based network was found to misclassify Acts 2 and 4 of *the Tempest*, and Acts 1 and 3 of *The Merry Wives of Windsor*, together with Acts 2 and 5 of *Valentinian*, and Act 4 of *Monsieur Thomas*, an overall success rate of 65%. As the probability of obtaining thirteen or more correct classifications by chance alone is 0.13, MNN's success is of only marginal significance.

The Horton-based network did considerably better, however, successfully classifying all but Acts 3 and 4 of *the Tempest* and Act 5 of *Valentinian*, a success rate of 85%; the results are shown in Table 2.

Although, as expected, both MNN and HNN were less successful when applied to acts rather than entire plays, the success rate of HNN was still very highly significant ($P < 0.001$). We thus conclude that both MNN and HNN are effective in discriminating authorship of entire plays, while HNN also remains effective down at the level of individual acts.

Table 2 Horton MLP results for core canon acts

Play	Horton SCM	Horton Verdict
Shakespeare		
Merry Wives of Windsor		
Act I	0.88	Shakespeare
II	0.74	Shakespeare
III	0.87	Shakespeare
IV	0.77	Shakespeare
V	0.93	Shakespeare
The Tempest		
Act I	0.91	Shakespeare
II	0.56	Shakespeare
III	0.31*	(Fletcher)
IV	0.37*	(Fletcher)
V	0.86	Shakespeare
Fletcher		
Monsieur Thomas		
Act I	0.29	Fletcher
II	0.30	Fletcher
III	0.29	Fletcher
IV	0.29	Fletcher
V	0.29	Fletcher
Valentinian		
Act I	0.30	Fletcher
II	0.30	Fletcher
III	0.29	Fletcher
IV	0.31	Fletcher
V	0.88*	(Shakespeare)

*Denotes apparent misclassification

5. Using the Networks on Disputed Works

Having investigated the relative powers of MNN and HNN to classify successfully both entire plays and individual acts, we applied each network to four works of particular interest: *The Double Falsehood, The London Prodigal, Henry VIII*, and *The Two Noble Kinsmen*.

All four plays have at some time been linked to Shakespeare and Fletcher. Although the anonymous *The Double Falsehood* has been associated with the Shakespeare apocrypha this play is now generally thought to be an adaptation of the now-lost *The History of Cardenio*, itself a collaboration between Shakespeare and Fletcher (Taylor, 1987). *The London Prodigal* is also anonymous and part of the Shakespeare apocrypha, but evidence supporting authorship by Fletcher has recently emerged from both stylometry (Merriam, 1992, Chapters 10 and 11) and sociolinguistic analysis (Hope, 1990).

Finally, interest in *Henry VIII* and *The Two Noble Kinsmen* stems from the fact that both have long been considered to be the product of collaboration between Shakespeare and Fletcher (Hart, 1934; Maxwell, 1962; Shoenbaum, 1967; Proudfoot, 1970).

Given this background, we applied both MNN and HNN to all four plays in their entirety, and then investigated the question of collaboration by applying HNN alone to individual acts of *Henry VIII* and *The Two Noble Kinsmen*. This produced the results shown in Table 3.

Table 3 Merriam and Horton MLP results for disputed plays

Play	Merriam SCM	Merriam Verdict	Horton SCM	Horton Verdict
Entire plays				
Double Falsehood	0.40	Fletcher	0.37	Fletcher
London Prodigal	0.31	Fletcher	0.30	Fletcher
Henry VIII	0.84	Shakespeare	0.94	Shakespeare
Two Noble Kinsmen	0.78	Shakespeare	0.65	Shakespeare
Plays by acts				
Double Falsehood				
Act I			0.66	Shakespeare
II			0.87	Shakespeare
III			0.29	Fletcher
IV			0.73	Shakespeare
V			0.29	Fletcher
London Prodigal				
Act I			0.89	Shakespeare
II			0.29	Fletcher
III			0.34	Fletcher
IV			0.28	Fletcher
V			0.30	Fletcher
Henry VIII				
Act I			0.98	Shakespeare
II			0.85	Shakespeare
III			0.97	Shakespeare
IV			1.00	Shakespeare
V			0.57	Shakespeare
Two Noble Kinsmen				
Act I			0.93	Shakespeare
II			0.30	Fletcher
III			0.32	Fletcher
IV			0.60	Shakespeare
V			0.91	Shakespeare

6. Analysis of Results

As Table 3 shows, both MNN and HNN agree that *The Double Falsehood* taken as an entire play is predominantly Fletcherian in style. Given this agreement of two different MLPs, and the more robust nature of results obtained when the MLPs are applied to entire plays, this finding appears to add evidential weight to the view that, despite being the product of an eighteenth-century adaptation, *The Double Falsehood* has considerable Fletcherian characteristics, agreeing with contemporary scholarship summed up by Metz (1989).

The SCMs for *The Double Falsehood* produced by both MNN and HNN are, however, somewhat higher than the ~ 0.3 value found by both MLPs for canon Fletcher works. This raises the possibility that the SCM is reflecting a Shakespearian influence on the play at the level of individual acts.

This possibility gains support from the application of HNN to individual acts of *The Double Falsehood*: we find three of the five acts have SCMs suggestive of a predominately Shakespearian influence. Given the greater statistical noise in the discriminators at the level of acts, less weight should be attached to these attributions, but they remain suggestive, none the less.

Similar remarks apply to the MLP findings with *The London Prodigal*: we find an overall Fletcherian attribution, but with some Shakespearian influence, especially in Act I. The results for *Henry VIII* taken as an entire play using both MNN and HNN indicate that it is predominately Shakespearian, a view that has long had its advocates (Foakes, 1957; Bevington, 1980). The SCM for the entire play is high, and even at the level of acts, all the attributions are to Shakespeare.

However, collaboration is not entirely ruled out: the relatively low SCM value for Act V suggests a strongly Fletcherian contribution to this part of *Henry VIII*, a view supported by Hoy (1956).

The results from both MNN and HNN for *The Two Noble Kinsmen* taken as an entire play also support an overall Shakespearian attribution, but the relatively low SCMs confirm current scholarly opinion of considerable collaboration between the two dramatists. The Horton-based network applied to individual acts provides more detailed information on this, attributing Acts I and V to Shakespeare, and Acts II and III to Fletcher. It also gives a relatively borderline SCM for Act IV, hinting at a considerable Fletcherian contribution to this act; all these assessments are in broad agreement with those of Proudfoot (1970) and Hoy (1956).

7. Conclusions

In this paper, we have set out the principles and practicalities of applying neural computation to stylometry. Multi-layer perception neural networks have two major advantages as a stylometric technique. First, experience gained by researchers in neural computation over a wide range of applications shows that MLPs are able to classify data even in the presence of considerable statistical noise. In addition, they are essentially non-linear classifiers, and can thus deal with interactions between stylometric discriminators, a feature denied traditional linear methods.

We have shown that after being trained using data drawn from 1,000-word samples taken from core canon works of Shakespeare and Fletcher, MLPs will successfully recognize known works of Fletcher and Shakespeare they have not encountered before.

In particular the MLPs were found to give excellent classification results when applied to entire plays, whose discriminator data are less subject to statistical noise. Furthermore, through the use of SCMs, they proved capable of reflecting authorship influence at the level of individual acts.

More specifically, when applied to disputed works the MLPs gave new evidential weight to the views of scholars concerning the authorship of four plays: *The Double Falsehood*, *The London Prodigal*, *Henry VIII* and *The Two Noble Kinsmen*. In the case of *The London Prodigal*, the evidence may now be sufficient to challenge the common assumption that, at 26, Fletcher was insufficiently mature to write such a play.

We believe that these results show that neural networks are a useful addition to current stylometric techniques. We cannot, however, overemphasize that—like any quantitative stylometric method—neural networks do not give incontrovertible classifications. Their true importance lies in their potential to provide an additional and independent source of evidential weight upon which literary scholars can draw.

We are ourselves now undertaking further research using MLP neural networks, and plan to report the results in due course (Merriam and Matthews, 1993).

Acknowledgements

It is a pleasure to thank Professor David Bounds of Aston University and Paul Gregory and Dr Les Ray of Recognition Research for their interest and advice, and for giving us access to their excellent NetBuilder software, without which this research may well have foundered. We also thank Dr Chris Bishop of AEA Technology, Dr Jason Kingdon of University College London for valuable discussions, and the anonymous referees whose constructive comments resulted in many improvements.

References

Anderson, J. A. and Rosenfeld, E. (eds) (1989). *Neurocomputing: Foundations of Research*, 4th printing. MIT Press, Cambridge.

Bevington, D. (ed.) (1980). *The Complete Works*. Scott, Foresman, Glenview.

Foakes, R. A. (ed.) (1957). *King Henry VIII* in *The Arden Shakespeare*. Methuen, London.

Hart, A., (1934). *Shakespeare and the Vocabulary of The Two Noble Kinsmen*. Melbourne University Press, Melbourne.

Hecht-Nielsen, R. (1990). *Neurocomputing*. Addison-Wesley, Reading.

Hope, J. (1990). Applied Historical Linguistics: Socio-historical Linguistic Evidence for the Authorship of Renaissance Plays, *Transactions of the Philological Society*, 88. 2: 201–26.

Horton, T. B. (1987). Doctoral thesis, University of Edinburgh.

Hoy, C. (1956). The Shares of Fletcher and His Collaborators in the Beaumont and Fletcher Canon (VII), *Studies in Bibliography*, 15: 129–46.

Maxwell, J. C. (ed.) (1962). *King Henry VIII*. Cambridge University Press, Cambridge.

Merriam, T. V. N. (1992). Doctoral thesis, University of London.

—— , and Matthews, R. A. J. (1993). Neural Computation in Stylometry II: An Application to the Works of Shakespeare and Marlowe, *Literary and Linguistic Computing* (submitted).

Metz, G. H. (ed.) (1989). *Sources of Four Plays Ascribed to Shakespeare*. University of Missouri Press, Columbia.

Minsky, M. and Papert, S. (1969). *Perceptrons*. MIT Press, Cambridge.

Proudfoot, G. R. (ed.) (1970). *The Two Noble Kinsmen* Edward Arnold, London.

Refenes, A. N., Azema-Barac, M., Chen, L., and Karoussos, S.A. (1993). Currency Exchange Rate Prediction and Neural Network Design Strategies *Neural Computing & Applications*, 1. 1: 46–58.

Rumelhart, D. E., and McClelland, J. L. (eds) (1986). *Parallel Distributed Processing* (I). MIT Press, Cambridge.

Shoenbaum, S. (ed.) (1967). *The Famous History of the Life of King Henry the Eighth*. The New American Library, New York.

Taylor, G. (1987). The Canon and Chronology of Shakespeare's Plays, *William Shakespeare: A Textual Companion*. Clarendon Press, Oxford.

Appendix

To encourage the greater use of neural networks in stylometry, the authors will happily provide .EXE files containing fully trained MLPs based on the Merriam and Horton discriminators to anyone sending a blank IBM-compatible 3.5" disk and return postage.