Using Simple Neural Networks to Correct Errors in Optical Data Transmission

Main Authors
Stephen Hunt, Yi Sun
Science and Technology Research Instititute,
University of Hertfordshire,
Hatfield, UK
e-mail: s.p.hunt@herts.ac.uk

Collaborating Authors 1
Neil Davey, Alex Shafarenko
Science and Technology Research Instititute,
University of Hertfordshire,
Hatfield, UK

Collaborating Authors 2
Sonia Boscolo, Sergei Turitsyn
Photonics Research Group,
School of Engineering and Applied Science,
Aston University,
Birmingham, UK

Abstract

We have demonstrated the applicability of neural-network-based systems to the problem of reducing the effects of signal distortion, and shown that such a system has the potential to reduce the bit-error-rate in the digitized version of the analogue electrical signal derived from an optical data stream by a substantial margin over existing techniques.

1 Introduction

Performance of fibre-optic communication links is typically affected by a complex combination of random processes (amplified spontaneous emission noise, polarization mode dispersion, and so on) and deterministic or quasi-deterministic effects (e.g. nonlinear inter- and intra-channel signal interactions, dispersive signal broadening and various cross-talks) that result from particular system design and operational regimes.

Any installed fibre link has its specific transmission impairments: its signature of how the transmitted signal is corrupted and distorted. This signature will change with data transmission speed, and may even change over time as the link's operating environment changes. Therefore there is great potential for the application of adaptive signal post-processing that can undo some of the signal distortions or separate line-specific distortions from non-recoverable errors. Signal post-processing in optical data communication can offer new margins in system performance in addition to other enabling techniques.

A variety of post-processing techniques have been already used to improve overall system performance, for example tunable dispersion compensation, electronic equalization and others (see [1, 3, 7 & 8] and references therein). Note that post-processing can be applied both in the optical and the electrical domain (after conversion of the optical signal into an electric current). Application of electronic signal processing for compensation of transmission impairments is an attractive technique that has become quite popular thanks to recent advances in high-speed electronics.

We have applied techniques of machine learning to adaptive signal post-processing in optical communication systems. We demonstrate the feasibility of bit-error-rate improvement by adaptive post-processing of the received electrical signal.

2 Background

At the receiver (typically after filtering) the optical signal is converted by a photodiode into an electric current. Detection of the digital signal requires discrimination of the logical ones and zeroes using a threshold decision of some sort. This can be done in different ways (e.g. by considering currents at certain optimized sample points within the bit time slots or by analyzing current integrated over some time interval) and is determined by the specific design of the receiver. The approach proposed in this paper and described in detail below is generic and can easily be adapted to any particular receiver design.

To improve system performance and minimize bit-error-rate, we propose here to use a trainable classifier so that the receiver may be adjusted to the characteristics of a particular line, and tuned over the lifetime of the line to take account of changes in operating conditions.

The classifier is used to identify whether a vector coding a sample of the signal waveform represents a 1 or a 0 in the digital domain, and tuning of the receiver is achieved by training the classifier with a set of waveform sample vectors for which the correct classifications are already known. This may be achieved by sending known bit sequences along the line to the receiver to provide examples of the transmission impairments that are specific to the line.

The classification problem addressed here is a special instance of the more general class of *time series* problems. Many of these problems have been addressed using supervised neural networks, typically employing a moving-window input vector. One special feature of the problem is the requirement for a simple two-way classification as the output of the classifier, suggesting that a single processing element may be used, as long as an appropriate activation function can be found. For the sake of simplicity (and processing speed) it would be useful if a linear activation function could be used; however, this requires the sets of input vectors to be linearly separable (or to have a sufficiently small amount of overlap that a system can be constructed that has an acceptably low mis-classification rate). In fact the problem at hand has several similarities to hardware branch prediction, a problem which is well known to be amenable to solution by the use of single-layered neural networks [2, 4 &5].

3 Description of the Data

The input data used in this work comprises a series of values, each of which quantifies a sample of the electric current obtained after converting an analogue optical signal into an electrical one. The analogue signal represents a sequence of bits. An analogue signal carrying a sequence of 5 consecutive bits is shown in Figure 1.

The analogue signal is sampled, with each sample being a floating-point number corresponding to the electric current at some point in time. There are 32 equally spaced sample points within a one bit time slot, so each bit in the signal is represented by 32 floating-point numbers.

We know the sequence of bits that was transmitted, so for each sequence of 32 samples in our data we know the bit that it represents. Therefore the data consists of a sequence of 32-ary vectors, each with a corresponding binary label.

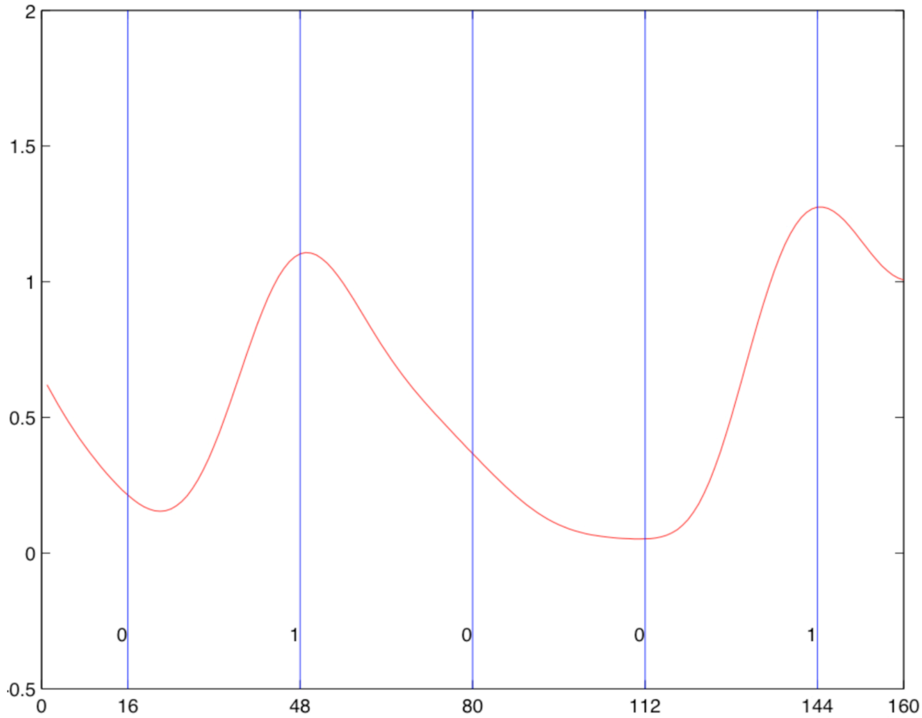

Figure. 1. An example of the analogue signal (representing light intensity) for a stream of 5 bits - 0 1 0 0 1

Some bit sequences give rise to signals that make the middle bit very hard to classify (Figure 2). In other cases both the central value and the cumulative energy of signal for the middle bit put it in one class when the human eye sees it as being in the other; such cases could be amenable to automatic identification. Figure 3 shows an example of a signal where the classification obvious to the eye but where a misclassification may occur. The central bit is a one but is misclassified as a zero from its energy alone.

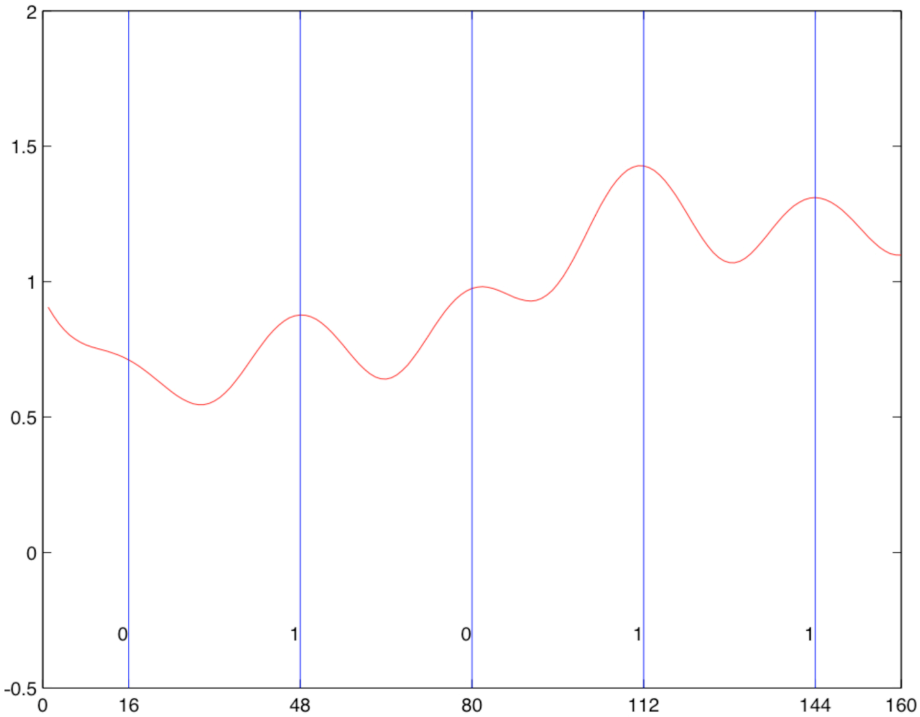

Figure 2. An example of an analogue signal where it is difficult to identify the middle bit correctly. It is meant to be a 0, but jitter has rendered it very hard to see.

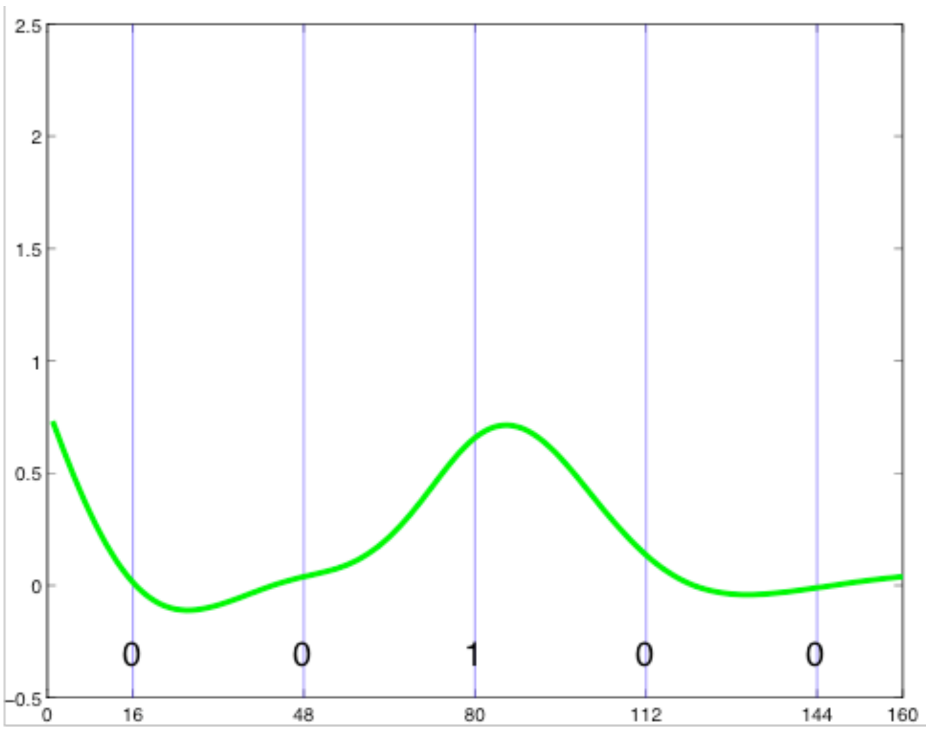

Figure 3. The central bit has been dragged down by the two zeros surrounding it and is classified as a zero from its cumulative energy. However to the human eye the presence of a 1 is obvious.

1.1 Representation of the Data

Different data sets can be produced depending on how the sampled electrical signal is represented. As well as representing a single bit as a 32-ary vector (which we have called the *Waveform-1* data set), it may be represented as a single cumulative energy value (the sum of the 32 values, *Energy-1*).

We may also want to take advantage of information that may be present in adjacent bits. To this end we have formed windowed inputs, in which the 3 vectors representing 3 contiguous bits are concatenated together with the label of the central bit being the target output (*Waveform-3*), see Figure 4. In addition, we have devised a representation that incorporate adjacent bit information by taking 3 consecutive energy values (*Energy-3*), and another that employs 5 consecutive energy values from a window of 5 bits, with 2 either side of the target bit (*Energy-5*). Table 1 gives a summary of all the different data sets.

Table 1. The different data sets used in the First Experiment.

Name	Arity	Description
Energy-1	1	The energy of the target bit
Energy-3	3	The energy of the target bit and one bit either side
Energy-5	5	The energy of the target bit and 2 bits either side.
Waveform-1	32	The waveform of the target bit
Waveform-3	96	The waveform of the target bit and the waveforms of the bits on either side

4 Classifiers Used

As already described, the classifiers need to be operationally very fast. Therefore the main classifier we use is a simple single layer neural network (SLN). Once trained (this is done off-line in advance, using *Iterated Re-weighted Least-Squares Training*, as described on pp.132-138 of [4]) an SLN can be built in hardware and function with great speed. For comparison purposes a classifier that employs an optimal energy threshold is implemented, where the threshold is found by 10-fold cross validation. A Support Vector Machine (SVM) with Gaussian kernel is also used for comparison purposes in one set of results.

5 Experiment 1

As indicated earlier it was thought that sequences of bits up to 2 on either side of the target bit could influence the categorization. There are 32 possible 5 bit patterns and in our data we have 900 examples of each. Each input vector was tagged according to which of these 32 different 5 bit patterns it came from, so that we could create training and test sets that contained representatives of all 32 patterns.

Each set of 900 vectors (one set for each of the 32 different 5 bit sequences) was randomly split into ten subsets of 90. We partitioned the 32 sets in turn, taking 810 vectors (9 of the 10 subsets of 90) for use in the training set, and the remaining 90 vectors for use in testing the generalization ability of the system. The overall training set for each run contained 25,920 (32 x 810) vectors, with the corresponding test set in each case being the 2,880 (32 x 90) vectors that were left out of the training set.

The data were partitioned in ten different ways to produce ten different training sets (and ten corresponding test sets). The results reported here are averages over the 10 different training and test sets. The main results are given in Table 2.

Table 2: The results of classifying the different test sets. Note: the Optimal Threshold result is over the full data set and is not an average, whereas the other results are averages of the 10 fold cross validation process and are therefore test sets of one tenth the size of the full data set.

Classifier	Data Set	Mean number of errors	Error Rate %
Optimal Threshold	*Energy-1*	179	0.622
SLN	*Energy-3*	8.35	0.29
SVM	*Energy-3*	5.76	0.20
SLN	*Energy-5*	8.64	0.30
SLN	*Waveform-1*	4.23	0.147
SLN	*Waveform-3*	6.91	0.24

The adaptable classifiers do give a significant improvement over the optimal energy threshold method, with the SLN using the *Waveform-1* dataset giving the best result. Interestingly the very simple classifier of the SLN/*Energy-3* combination more than halved the error rate when compared to the optimal threshold method. This classifier is simply a single unit with 3 weighted inputs. There is no evidence here that the information in the two bits either side of the target bit (*Energy-5*) is useful.

One difficulty for the trainable classifier is that in this data set the vast majority of examples are straightforward to classify. The hard cases are very sparsely represented, so that, in an unusual sense, the data is imbalanced. To examine more closely how a trainable classifier performs on the difficult cases Experiment 2 was undertaken.

6 Experiment 2

As described above, of the 28,800 bits in the data stream all but 179 are correctly identified by an energy threshold. Out of the 32 distinct sequences of 5 bits identified for experiment 1, only 10 are represented in this 'difficult' subset. These ten sequences are:

0 0 1 0 0
0 0 1 0 1
0 1 0 1 1
1 0 1 0 0
1 0 1 0 1
1 0 1 1 0
1 0 1 1 1
1 1 0 1 0
1 1 0 1 1
1 1 1 0 1

The majority of these involve a 1 0 1 or 0 1 0 sequence around the middle bit, and these are the patterns for which difficulties are most likely to occur. In this experiment we attempt to concentrate on the learning of vectors that are likely to be hard to classify because they are drawn from these 'difficult' sequences

We selected 91 of the 179 vectors that were mis-classified by an energy threshold classifier for inclusion in the training set. These included roughly half of the examples from each of the ten sequences shown above. A further 85 vectors were used for testing; the 3 remaining mis-classified vectors were not used as each was a singleton in its class. Each of the 91 mis-classified vectors was included in the training set 20 times, with the rest of the training set being made up of 100 examples of each of the ten 'difficult' sequences that were correctly identified by the energy threshold. This gives a training set of 2820 vectors. An SLN was then trained using these vectors and the resulting classifier was tested on the remaining 85 test patterns. The results are given in Table 3 below.

Table 3. The results of classifying the 85 patterns in the 'difficult' test set when the SLN is trained using a set containing a high proportion of 'difficult' vectors.

Data Set	Number of misclassifications	Error Rate %
Energy-1	67	78.7
Energy-3	8	16.5
Waveform-1	4	8.24
Waveform-3	5	11.76

Figure 4 shows the input profiles of a number of examples of one of the 'difficult' 5 bit sequences (10100). The job of the classifier is to identify the central bit as a '1'. It is clear that the SLN outperforms the energy thresholding system.

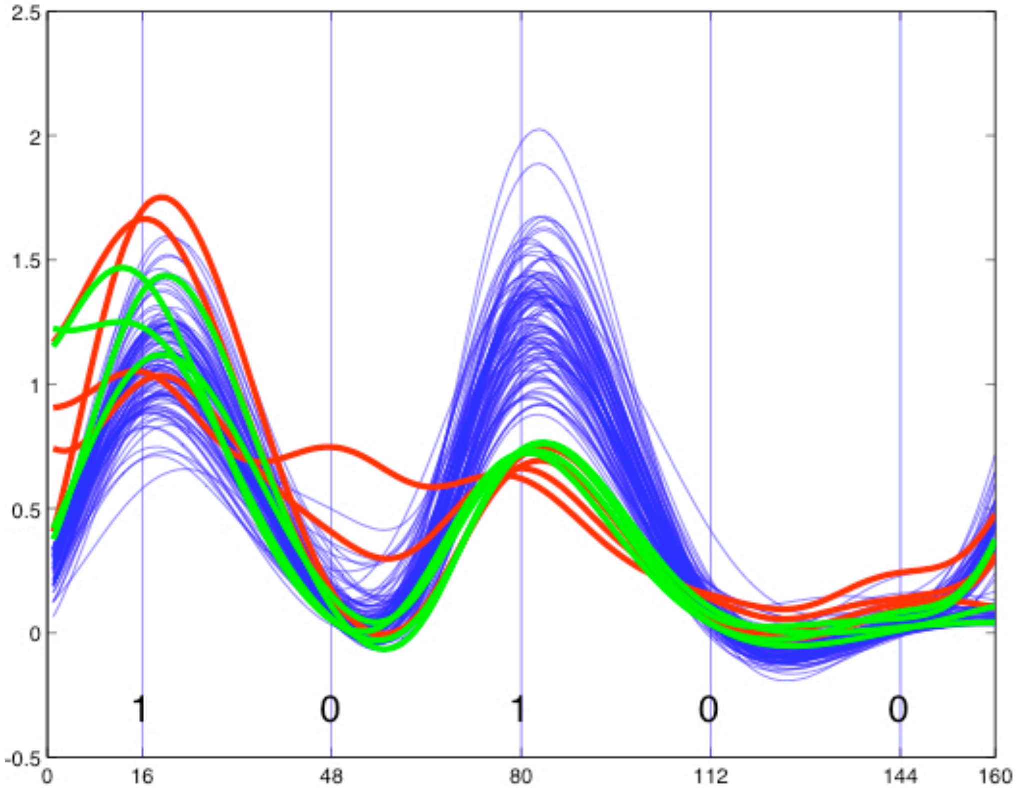

Figure 4. The blue lines represent those input streams that are correctly classified by both SLN and energy threshold systems. The green lines represent those that are correctly classified by the SLN but mis-classified by the energy threshold system, and the red lines are those that are misclassified by both.

With only 4 patterns misclassified by the best SLN classifier the overall error rate for the entire data set is now well below the desired 1 in a thousand. It is also interesting that the Energy-3 representation gives only 8 misclassifications. This shows that much useful information is present in the two bits either side of the target bit.

Of course the classifier presented in this Section is not a workable solution. It is not known a-priori whether a particular signal is part of the difficult set or should simply be sent to the energy thresholder. What the results do demonstrate, however, is that given enough balanced examples it is possible to correctly identify many of the difficult patterns.

7 Discussion

The fast decoding of a stream of data represented as pulses of light is a commercially important and challenging problem. Even a small reduction in bit error rates can lead to a useful increase in data throughput.

Computationally, the challenge is in the speed of the classifier and the need for simple pre- and post-processing. We have therefore restricted our classifier to be, for the most part, a single layer network and the data is either a sequence of electric current values or just the total energy of the pulse over a specified time. Experiment 1 showed that by using an SLN trained with a set of 32-ary samples of the input stream the bit error rate could be reduced from the 0.62% achieved by the existing energy threshold method to 0.24%.

This figure is still quite high and we suspected that the explanation might be that despite the data set being very large (28,800 examples) the number of difficult examples (those mis-classified by the threshold method) was very small and the behaviour of the system was dominated by the number of straightforward examples. To see if we could correctly identify a significant number of these infrequent but difficult examples we undertook experiment 2. Here it was shown that in a dataset in which these difficult examples were in the majority it was possible to train an SLN to correctly identify almost 90% of a set of unseen examples which were themselves mis-classified by the energy threshold method.

Preliminary results from a third experiment (not presented here) suggest that both the sampling frequency and the sample resolution may be reduced with little or no deleterious effect on the SLN's ability to correctly classify analogue signals representing 'difficult' bit sequences. This may turn out to be important, as both the sampling rate and the sample resolution will influence the cost of the hardware required to do the job in real time. However, in the absence of results from large-scale tests on real-world data it would be premature to make any recommendations regarding an appropriate sampling rate or sample resolution for use in a working system.

In future we will be training and testing the SLN system with datasets that include a higher proportion of difficult bit sequences, and with large-scale datasets derived from communications channels that have much higher levels of noise than those investigated so far.

References

[1] Bulow, H. (2002) Electronic equalization of transmission impairments In *OFC 2002*, Anaheim, CA, Paper TuE4.

[2] Calder, B., Grunwald, D., Jones, M., Lindsay, D., Martin, J., Mozer, M. & Zorn, B. (1997) *Evidence-based Static Branch Prediction using Machine Learning*, ACM Transactions on Programming Languages and Systems, **19**, 1997

[3] Haunstein, H.F. & Urbansky, R. (2004) Application of Electronic Equalization and Error Correction in Lightwave Systems, In *Proceedings of the 30th European Conference on Optical Communications (ECOC)*, Stockholm, Sweden, September 2004

[4] Hunt, S.P. Egan, C. & Shafarenko, A. (2003) *A Simple Yet Accurate Neural Branch Predictor*, *Artificial Intelligence and Applications, AIA 2003*, ACTA Press, Malaga, Spain, 2003.

[5] Jiménez, D.A. & Lin, C. (2002) *Neural Methods for Dynamic Branch Prediction*, ACM Transactions on Computer Systems, **20**, 369-397, 2002

[6] Nabney, I.T. (2003) *NetLab: Algorithms for Pattern Recognition*

[7] Rosenkranz, W. & Xia, C. (2007) Electrical equalization for advanced optical communication systems, *AEU - International Journal of Electronics and Communications*, **61** (3), March 2007, 153-157

[8] Watts, P.M., Mikhailov, V., Savory, S. Bayvel, P. Glick, M., Lobel, Christensen, M.B., Kirkpatrick, P., Shang, S. & Killey, R.I. (2005) Performance of single-mode fiber links using electronic feed-forward and decision feedback equalizers, *IEEE Photon. Technol. Lett.* **17** (10), 2206 – 2208, Oct. 2005