Connectionist Learning Using an Optical Thin-Film Model

Martin Purvis and Xiaodong Li
Computer and Information Science
University of Otago
Dunedin, New Zealand

Abstract

An alternative connectionist architecture to the one based on the neuroanatomy of biological organisms is described. The proposed architecture is based on an optical thin-film multilayer model, with the thicknesses of thin-film layers serving as adjustable 'weights' for the computation. The nature of the model and some example calculations that exhibit behaviour typical of conventional connectionist architectures are discussed.

1 Introduction

Although connectionism has taken its inspiration from the neuroanatomy of biological organisms, the essence of this form of computation does not require strict adherence to the biological exemplar: architectures that differ significantly from the biological model may exhibit interesting computational properties. In this paper we describe an alternative architecture, inspired by the technology of optical thin film multilayers and first reported in [1], that exhibits connectionist learning capabilities. We first provide a brief description of the architecture in section 2 and then describein section 3 some example model calculations that compare with conventional neural network computations.

2 The Optical Thin Film Architecture

When a light beam is incident on a single thin-film, it undergoes multiple reflections and refractions, the various component beams of which interfere with each other to produce an overall reflectance and transmittance governed by the refractive index and thickness of the film and the wavelength of the incident light. When several thin-films of differing materials are deposited on top of each other to form a thin-film multilayer (Figure 1), there are multiple reflections and transmissions at each boundary, and it is possible to derive a recursive algebraic expression for the overall reflection coefficient of the resulting multiple beam interference.

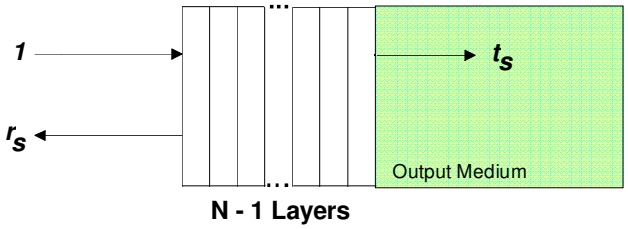

Figure 1. N - 1 thin-films deposited on a substrate.

Suppose an additional thin-film layer with complex refractive index $\boldsymbol{n}_N = n_N + ik_N$ and thickness d_N is brought up to the existing multilayer structure shown in Figure 1. (The imaginary component of the complex refractive index accounts for any absorption of the thin-film material.) It can be shown [2] that the combined complex reflection coefficient, r_N, of the resulting N layer system comprising the structure and the added layer is

$$r_N = \frac{f_N(1 - \varepsilon_N^2) + (\varepsilon_N^2 - f_N^2)r_{N-1}}{(1 - f_N^2\varepsilon_N^2) + f_N(1 - \varepsilon_N^2)r_{N-1}} \qquad \textbf{(1)}$$

where

$$f_N = \frac{(\hat{n}_N - 1)}{(\hat{n}_N + 1)} \qquad \tau_N = \frac{2}{(\hat{n}_N + 1)}$$

$$\varepsilon_N = \exp(i2\pi\hat{n}_N d_N/\lambda)$$

The factor ε_N accounts for both phase shift and amplitude attenuation due to light propagation.

The optical reflectance value that is typically measured is obtained by taking the square of the absolute magnitude of the reflection coefficient given by (1). Almost any desired spectral characteristic of the reflection can be obtained by choosing a sufficient number of layers with appropriate values of n and d for each thin-film layer.

For our computational purposes, we may view the reflection coefficient with real and imaginary components always in the range of -1.0 to 1.0 for each value of the optical wavelength as the output. Input is encoded into the thin-film system by adding a small, scaled offset to the refractive index for a given layer. Typically, the *i-th* input value is scaled and added as an offset to the base index n_{Bi} of the *i-th* thin-film layer of a multilayer structure so that $n_i = n_{Bi} + n_{\Delta i}$. Thus the layer refractive indices for this system are considered to be variable about some average values.

Training involves adjusting the individual layer thicknesses and then calculating the multilayer reflection coefficient (or, alternatively, the reflectance) for each optical wavelength for which output is specified. The resulting spectral output $r(\lambda)$ is compared to a target output $r_T(\lambda)$ for each wavelength and a merit function *M* is used to determine the error for the given configuration:

$$M = \sum_{\lambda} |r(\lambda) - r_T(\lambda)|^2 \qquad \textbf{(2)}$$

or

$$M = \sum_{\lambda} |R(\lambda) - R_T(\lambda)| \qquad \textbf{(3)}$$

Training is continued until a configuration of layer thicknesses is found that produces a satisfactory value of the merit function.

Calculation of the reflection coefficients (as simulated on a computer) is computationally expensive for many-layered thin-film structures, so rather than an exhaustive examination of the full range of thickness combinations, we have adopted the "N-squared scan" approach of Heavens and Liddell [3]. This approach yields a pruned search space so that for a N-layered structure with *q* different thicknesses to be considered for each layer, there are only qN^2 film stack evaluations, rather than the q^N evaluations that would be required by the exhaustive approach. Experience has shown that this reduced search-space can still yield acceptable results [3,4]. Additional details concerning our use of the N-squared scan approach other algorithmic considerations that are used to improve the efficiency of the computation are given in [1].

3 Computational Learning with the Model

We give two example calculations with the optical thin-film model that have also been used to demonstrate the properties of conventional feed-forward neural network architectures. The first involves pattern recognition of roman alphabetic characters presented by means of a 5 by 5 grid of pixels. The second example involves the discrimination of data for varieties of iris plant into three recognized species [5].

3.1 Pattern recognition of alphabetic characters.

A computational model of an optical thin-film multilayer stack was trained with the 5 by 5 pixel grid patterns representing 'I', 'O', 'C', and 'X' shown in Figure 2. The stack was specified to have 25 layers with values of the base refractive index n_{Bi} for each layer alternating between 1.2 for the odd-number layers and 4.0 for the even-number layers. The inputs were encoded as small values to be added to n_{Bi} -- either 0.0 or 0.4 depending on whether a pixel in the 5 by 5 grid is blank or shaded.

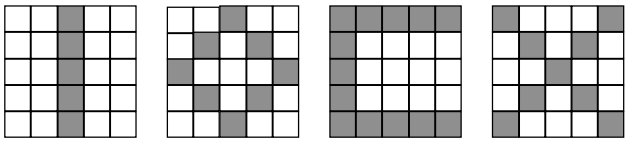

Figure 2. Four 5 by 5 pixel patterns used for training.

Target reflectances for each pattern at 8 different wavelengths ranging from 10.0 μm to 12.8 μm were selected. (For all calculations reported in this paper and in the figures, optical wavelengths and thin-film layer thicknesses are expressed in terms of micrometers. The range of values selected correspond values for optical thin-film multilayers in the microwave region.) The target reflectance values for the four 5 by 5 grid patterns are shown in Figure 3.

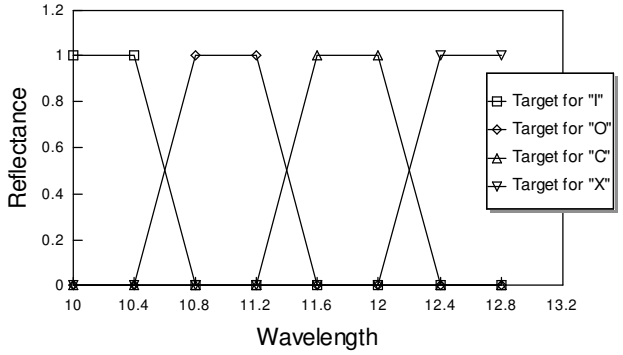

Figure 3. Target reflectances for "I","O","C" and "X".

The target spectral response for the character 'I' has a peak value of 1.0 in the wavelength range of 10.0 to 10.4, descends to 0.0 at a wavelength of 10.8, and then remains steady at 0.0 out to the wavelength of 12.8. The target response for the character 'O' has peak values in the wavelength range of 10.8 to 11.2 and is zero for wavelengths below 10.4 and above 11.6. The target response for the character 'C' has peak values for wavelengths from 11.6 to 12.0, with zero values below a wavelength of 11.2 and above a wavelength of 12.4. The target response for the character 'X' has peak values for wavelengths between 12.4 and 12.8, with zero values for wavelengths below 12.0.

After the optical thin-film stack was trained, its spectral behaviour when presented with the four training set grid patterns shown in Figure 4.

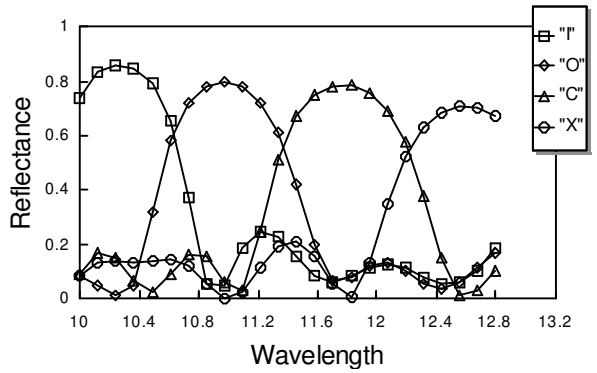

Figure 4. Training result for the 4 patterns.

As can be seen from Figure 4, the trained thin-film stack approximates the target curves (although with rounded shoulders).

After training, the thin-film stack was tested for its ability to recognize the bit-patterns in the presence of noise. Four "noisy" versions of the letter 'I' were used for this purpose, and their bit patterns are shown in Figure 5.

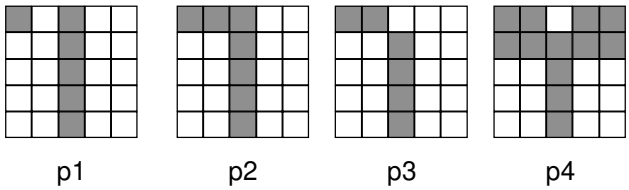

Figure 5. Four "noisy" versions of the letter 'I'.

The first pattern, p1, shown in Figure 5 has only one pixel different from the original pattern for 'I' shown in Figure 2. The patterns p2, p3, and p4 are successively more deviant from the original 'I' grid pattern. The spectral response of the trained thin-film stack to the letter 'I' and the four noisy patterns is shown in Figure 6.

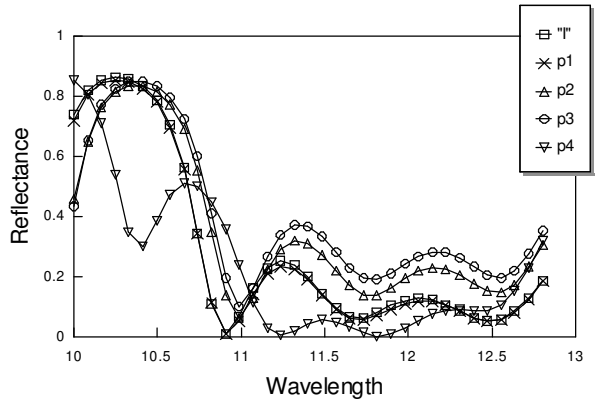

Figure 6. Test result on 'I' and 4 noisy versions.

The four response curves to the noisy patterns p1, p2, p3, and p4 shown in Figure 6 show successively greater deviation from the target curve 'I' shown in Figure 3. In particular the response curve for p4 is considerably different from the target for 'I'. Each of the four noisy patterns have successively larger values of merit function error evaluations (using equation (3) above). By setting an appropriate tolerance value for merit function evaluations, we can tune the trained thin-film stack to accept some noisy patterns versions of the letter 'I' and reject patterns that fall outside the tolerance range.

Thus the optical thin-film model can be trained to recognize certain bit-patterns and can respond correctly even in the presence of noise, just as is the case with conventional neural network architectures.

3.2 Classification of iris data.

The iris data set [5] has been frequently used as an example for discriminant analysis of real data. The dataset was collected for three species of iris and comprised four measurements (petal length, petal width, sepal length, and sepal width) for 50 samples of each species. One of the species is thought to have arisen as a hybrid species and displays a mixture of features of the other two species, making discrimination among the species by means of the measured properties somewhat difficult [6]. One class is linearly separable from the other two, while the other two are not linearly separable from each other.

For each iris data example used, e.g. {5.1, 3.5, 1.4, 0.2}, it is necessary to scale the values appropriately for encoding into the thin-film system. On this occasion the output used was the complex reflection coefficient (rather than the single-valued reflectance used in section 3.1), and a thin-film stack of only four layers was selected with values of the base refractive index for each n_{Bi} alternating between 1.2 and 4.1. The training was carried out for the optical wavelength values: 5.0, 5.3, and 5.6 μm, and the target reflectance coefficient values used are shown in Table 1.

Table 1. Target reflection coefficients

λ	Class 1		Class 2		Class 3	
	r_{RE}	r_{IM}	r_{RE}	r_{IM}	r_{RE}	r_{IM}
5.0	0.2	0.2	0.3	0.3	0.4	0.4
5.3	-0.5	0.1	-0.4	0.3	-0.3	0.4
5.6	0.1	-0.5	0.3	-0.4	0.4	-0.3

The thin-film stack was then trained, using 120 examples (40 randomly chosen from each of the three classes). Training had the effect of moving the response reflection coefficient points as close as possible to the 3 target points for each optical wavelength value in an effort to separate the collection of output reflection coefficient values as much as possible. Although data from class 1 were completely separated from

the other two classes, there were a couple of points from class 2 and class 3 that always overlapped into each other's regions.

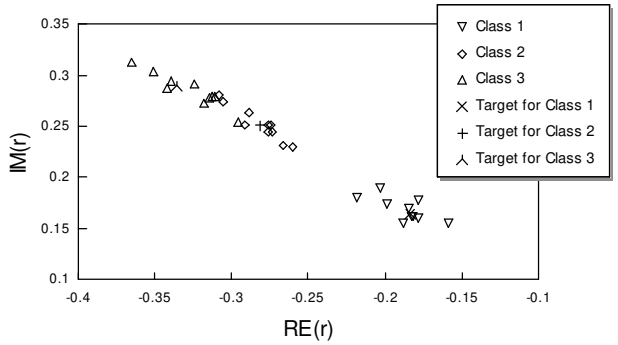

Figure 7. Test results on 30 iris data samples at λ = 5.0 μm.

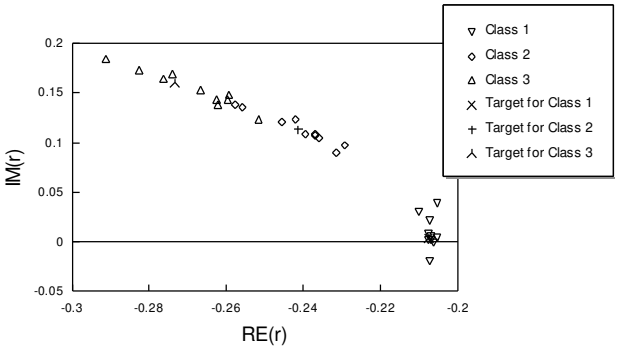

Figure 8. Test results on 30 iris data samples at λ = 5.3 μm.

After this training, a "winner-take-all" approach [7] was used to set points in the output range to be used for discrimination of samples (during test runs) into the three species. This was accomplished by finding the average values of all the output points at each optical wavelength and for each training class. Thus each of the three species will have a set of characteristic values for output reflection coefficients.

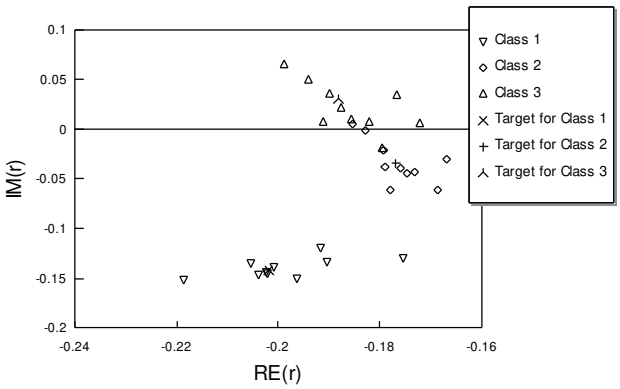

Figure 9. Test results on 30 iris data samples at λ = 5.6 μm.

When test example is entered into the system, it is a matter of calculating the output reflection coefficient for the test example (at each wavelength) and then using the merit function evaluation of equation (2) to determine to which species class the example output is closest.

Output for thirty novel examples is shown in Figures 7, 8, and 9. Of the thirty examples only 2 were misclassified by the thin-film system. When the system was given all 150 samples, 5 were classified incorrectly.

4 Conclusions

The optical thin-film architectural model discussed here has been shown to exhibit the kind of connectionist learning typical of conventional neural network architectures. Results of further experiments, including the XOR and 4-bit parity problem, and more detailed discussion will be published by the authors in a forthcoming paper.

An optical realization of this architecture would offer interesting possibilities with respect to computational efficiency, although such a realization has yet to be carried out to our knowledge. Some further considerations on this matter are provided in [1].

References

[1] Purvis, M. K. and Xiaodong, L., "Connectionist Computations Based on an Optical Thin-Film Model", *Proceedings of the First New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems,* IEEE Computer Society Press, Los Alamitos, California, 1993, pp. 130-133.

[2] Case, W. E., "New Synthesis Method for Optical Thin-film Coatings", *Applied Optics*, 22:24 (1983), pp. 4111-4117.

[3] Liddell, H. M., *Computer-aided Techniques for the Design of Multilayer Filters*, (1981), Adam Hilger, Bristol.

[4] MacLeod, H. A., Case, W. E., and Purvis, M. K., "Design of a Vanadium Dioxide Switchable Coating", in *Topical Meeting on Optical Interference Coatings*, Opt. Soc. of Am., Monterrey, CA, 1984.

[5] Anderson, E., "The Irises of the Gaspe Peninsula", *Bulletin of the American Iris Society,* vol 59, pp. 2-5.

[6] Ripley, B. D., "Statistical Aspects of Neural Networks" in *Networks and Chaos -- Statistical and Probabilistic Aspects*, Chapman and Hall, London, pp. 40-123.

[7] Hart, A., "Using Neural Networks for Classification Tasks – Some Experiments on Datasets and Practical Advice", *J. Opl. Res. Soc.,* vol. 43, pp. 215-226.