Interferometer-based Photonic Circuit Classifier Showing >90% Accuracy for Well-known Iris Dataset without Utilizing Nonlinear Activation Function

Guangwei Cong*, Noritsugu Yamamoto, Takashi Inoue, Yuriko Maegami, Morifumi Ohno, Makoto Okano, Shu Namiki, and Koji Yamada

Electronics and Photonics Research Institute, National Institute of Advanced Industrial Science and Technology, Onogawa 1-16, Tsukuba, Ibaraki 305-8569, Japan

**gw-cong@aist.go.jp*

Abstract: We demonstrate that interferometer-based photonic circuits can perform classification by only phase control even without activation functions, which can classify well-known Iris dataset with >90% accuracy in simulation, showing simple photonic implementation for machine learning.

OCIS codes: (250.5300) Photonic integrated circuits; (130.3120) Integrated optics devices

1. Introduction

Classifier is one of the important categories in machine learning, which has many applications in medical diagnosis, market predictions, plant classifications, etc. Usually, it can be realized by building deep neural networks [1] or other algorithms such as decision tree and support vector machine. Modern DNNs are able to deliver superior performances than other algorithms on a variety of talks by employing a deep hierarchy of layers and various activation functions, but at the cost of very high computational complexity. Recently, implementation of DNN in integrated photonics could greatly save computational resources and time since calculation in neural networks can be emulated by optical propagation inside the photonic circuits [2], activating the study of photonic neural network [3]. However, the nonlinear activation functions were still necessary to be simulated by computer program, which limits to achieve fully optical DNN without interfering with microprocessor once the training is ready. Even though integration with OEO conversion devices, SOA, or saturation absorbers offers hybrid solution to realize activation functions, scalability would be decreased. In this work, we demonstrate that without necessity to build full DNN with nonlinear activation functions, a well-trained photonic interferometer can perform functions of classifier such as plant classifications. This photonic circuit is consisted of only splitters and Mach-Zehnder interferometers (MZIs) and accepts optical phase control only. There is also only one light input port and all parameters are input through optical phase as well. After being trained, just by light propagation and detection, this photonic interferometer can perform the classification functions that are usually done by complex algorithms or DNN. This photonic interferometer is different from the traditional photonic DNN since it has no specific layer of any nonlinear activation functions, but it can be regarded as a modified neural network consisting of several layers of MZIs, which is in essence the coherent matrix multiplication, instead of linear matrix multiplication in traditional neural networks. Even though it cannot perform some functions of DNN such as image recognition, it works well for classification functions, for which therefore it is not necessary to build DNN and such a simple photonic circuit can greatly mitigate the energy and time to implement neural networks for classifier applications. In this study, we demonstrate how to implement photonic classifier to recognize the well-known Iris dataset that is widely used in machine learning [4] and report >90% high recognition accuracy that is comparable to that obtained by DNN [5]. This scenario of photonic classifier shows a much scalable and simple photonic implementation for machine learning.

2. Iris dataset and photonic classifier

Iris dataset is famous for testing how the algorithms and DNN models are efficient to classify Iris plant species according to leaf attributes [4]. The dataset has 150 instances (*i.e.*, samples) in total and Fig. 1(a) shows a screen shot for some of them. As seen, the dataset consists of four leaf attributes (sepal length and width, petal length and width) and three corresponding classes (Setosa, Versicolor, Virginica). Each class has 50 instances and each instance is described by one group of attributes and has a corresponding species. This work is to use a simple photonic classifier to classify this dataset. The structure of photonic classifier is shown in Fig. 1(b). There is only one light input port which is spitted equally or non-equally into four paths. Connecting these four paths are a column of MZIs that are used for inputting attribute parameters $\langle \boldsymbol{x} \rangle$ which are normalized phase parameters obtained by normalizing original parameters in Fig. 1(a). Details of normalization will be explained in next section. A universal photonic circuit of 4×4 Clement topology follows this column of MZIs, which is consisted of two sets of MZI matrix (25 MZIs in total). This topology was explained in Ref. [6]. The phase difference $\langle \boldsymbol{\phi} \rangle$ in MZIs of this photonic circuit will be trained. At the output, we can sign any three ports of either three from four MZIs to mark the

answer of Iris class. As seen in Fig. 1(b), ports 1, 3, and 5 are assigned to stand for Setosa, Versicolor, and Virginica, respectively. By monitoring the optical power, the port id (i) of maximum power (P_i^{max}) gives the answer of Iris class and the recognition is wrong if the maximum power occurs at other ports but not the right one when inputting corresponding attribute parameters for a species. Training is to find one unique phase distribution of $<\boldsymbol{\phi}>$ to make images to appear at the right image positions for as many instance parameters $<\boldsymbol{x}>$ in the dataset as possible.

No.	Sepal length	Sepal width	Petal length	Petal width	Species
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3	1.4	0.2	Setosa
3	4.7	3.2	1.3	0.2	Setosa
:	:	:	:	:	:
51	7	3.2	4.7	1.4	Versicolor
52	6.4	3.2	4.5	1.5	Versicolor
53	6.9	3.1	4.9	1.5	Versicolor
:	:	:	:	:	:
101	6.3	3.3	6	2.5	Virginica
102	5.8	2.7	5.1	1.9	Virginica
103	7.1	3	5.9	2.1	Virginica
:	:	:	:	:	:

(a) Unit: cm

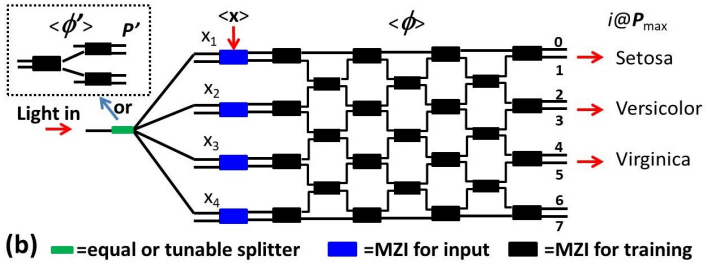

Fig. 1. (a) Iris dataset. (2) Interferometer-based photonic circuit classifier for 4-attribute inputs. The splitter can be equal or tunable one.

The photonic circuit was simulated by transfer matrix method with a phase distribution $<\boldsymbol{\phi}>$ and trained by bacterial foraging algorithm (BFA) which can be referred to in our previous works [7,8]. The whole device can be described as the output power vector $\boldsymbol{y} = \prod f'(\boldsymbol{\phi}) \cdot f(\boldsymbol{x})$, where f and f' are matrices of Mach-Zehnder transfer functions, presenting coherent matrix multiplication instead of linear matrix multiplication. This interferometer-based photonic classifier only accepts phase control for both input parameter $<\boldsymbol{x}>$ and training parameter $<\boldsymbol{\phi}>$. In an actual device, they will be voltage signals and all MZIs could be the thermo-optic ones as used in matrix switches [9]. This device can be fabricated on silicon photonic platform and the scalability depends on the splitter. If a 1×32 splitter can be fabricated, this classifier can perform classification for 32 attributes. The splitter can be made tunable, instead of equal splitting, by utilizing cascaded MZI with phase distribution $<\boldsymbol{\phi}'>$, as shown in the inset of Fig. 1(b), providing another freedom for training. If photodetectors are on-chip integrated, there is only one input fiber so that fiber packaging will be simpler than previous multi-port switches [9].

3. Training and classification accuracy

The first step before training $<\boldsymbol{\phi}>$ is to determine $<\boldsymbol{x}>$ by normalizing the original attribute values ($\boldsymbol{v}$) in dataset of Fig. 1(a). The normalization is done by $x_{jk}=\alpha\cdot[v_{jk}-min(\boldsymbol{v})]/[max(\boldsymbol{v})-min(\boldsymbol{v})]$ where j (=1~4) and k (=1~150) denote attribute and instance number, respectively, and $max(\boldsymbol{v})$ and $min(\boldsymbol{v})$ mean the maximum and minimum values among all attributes, respectively. The coefficient α is the target normalization value. In simulation, it was set to be 2π in the first place and we consider pre-power tuning by adopting tunable splitters in the inset of Fig. 1(b). The influences of normalization values and whether pre-power tuning is used will be discussed in the later part. Then, the classifier is trained by BFA [7,8] to search solutions of $<\boldsymbol{\phi}>$ and $<\boldsymbol{\phi}'>$ (for pre-power tuning) to achieve $max[P_i(\boldsymbol{\phi}_{jk},\boldsymbol{x}_{jk})]$ at the port i = 1 for Setosa (k = 1–50), i = 3 for Versicolor (k = 51–100), and i = 5 for Virginica (k = 101–150). The searching range of $\boldsymbol{\phi}$ is $[-\pi, \pi]$ and the minimum step is 0.001π. For real implementation in a thermo-optic device, all these parameters can be converted to physical equivalents; α will be double π-phift power ($2P_\pi$) and then x_{ij} can be converted to voltage by rooting $2P_\pi\cdot[v_{jk}-min(\boldsymbol{v})]/[max(\boldsymbol{v})-min(\boldsymbol{v})]$. The range can be set to $[0, 2P_\pi]$ and the minimum step will be about 1 mV if V_π = 3V. To mention that this work reports full classification in the first step (all samples were used for training) and dividing samples into training and testing will be reported in near future.

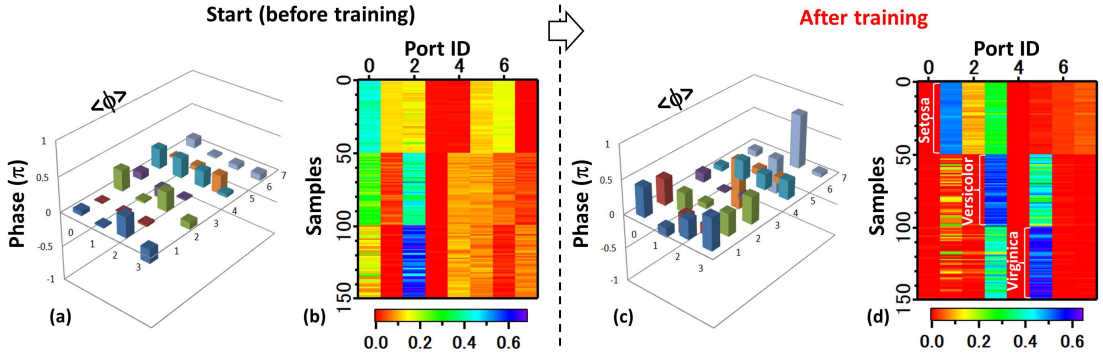

Fig. 2. (a) Phase landscape in the interferometer initialized randomly within $[-0.3\pi, 0.3\pi]$ at the start (before training) and (b) its corresponding initial optical power mapping at every output port ID for all 150 Iris samples. (c) Phase landscape and (d) optical power mapping after training indicating the clear classification of three kinds of Iris. Input optical power is one which is the reference of power mapping in (b) and (d).

The initial phase states of all MZIs are considered not homogenous with random errors within ±0.3π (equivalent to ~80nm length difference of silicon waveguide). For example, Fig. 2(a) shows the initial state of $<\phi>$, upon which the output power mapping is given in Fig. 2(b) for all ports and all samples. As seen, the maximum-power port ID for each sample does not appear at the correct position. However, after training, the solution of $<\phi>$ in the interferometer is shown in Fig. 2(c) under which we can observe three clear color regions of maximum-power port IDs in the output power mapping, as seen in Fig. 2(d), indicating correct classification of three Iris types. To evaluate the accuracy, we need examine the port position of maximum power for every sample. Ideally in absolute comparison, we can always find a maximum-power port despite of a small difference between its power and others. But in reality, we should consider how distinguishable they are; thus, we define a relative ratio r in judging the maximum-power port ID by satisfying $P_i^{\max} - P_j > rP_i^{\max}$, where $P_i^{\max}$ is the maximum power observed at the port i and P_j is the power at other ports j. r=0 means absolution comparison and we also consider r=0.05 and 0.1 which mean that the port ID is correct only if its power is larger than others by 5% and 10%, respectively. For the trained result in Fig. 2(d), the accuracy (percent of correct port numbers) is larger than 90% for all three ratios, as seen in Fig. 3(a) which shows the convergence progress of accuracy in relation of check times (one check time had the optical propagation loops of 17~20 times case by case, indicating $7\sim8\times10^4$ loops in total). For r=0, the accuracy is larger than 97%, showing the principle feasibility of photonic classifier. For r=0.05 and 0.1, it is larger than 90%, indicating the viability since 5~10% contrast is enough for most photodetectors to judge the maximum power. Such training was repeated more for different random initial phase states and even though the convergence progress may vary due to stochastic nature as seen in Fig. 3(b), >90% accuracy is guaranteed for r=0 and 0.05. For r=0.1, the accuracy ranges from 86% to 92%, comparable to that of DNN [5]. Finally, we examine the effects of normalization and pre-power tuning on the accuracy. Figs. 3(c) and 3(d) compare the convergence of accuracy between two normalization targets (π and 2π) without and with using pre-power tuning, respectively. Normalization to 2π is better than π in obtaining higher accuracy and faster convergence speed no matter whether pre-power tuning is adopted. For the same normalization under any ratio r, pre-power tuning can offer better classification accuracy.

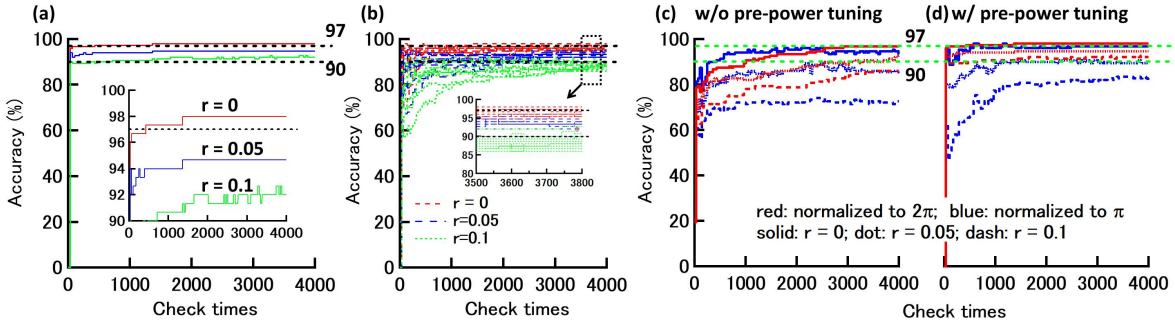

Fig. 3. (a) Convergence progress of classification accuracy for Fig. 2(d) and (b) repeated training results of accuracy for different initial phase states. (a) and (b) use pre-power tuning and 2π as the normalized target. Accuracy comparisons on two normalization targets (π and 2π): (c) without (w/o) pre-power tuning and (d) with (w/) pre-power tuning through a tunable splitter.

4. Summary

An interferometer-based photonic classifier was proposed and demonstrated in simulation, which could perform classification without using activation functions. It can classify well-known Iris dataset with >90% accuracy even considering a 5~10% contrast ratio, showing simple photonic implementation for machine learning applications.

Acknowledgment

This work was partly supported by JST CREST Grant Number JPMJCR15N4, Japan

[1] Y. LeCun *et al.*, "Deep learning," Nature **521**, 436 (2015).
[2] Y. Shen *et al*, "Deep learning with coherent nanophotonic circuits," Nature Photonics **11**, 441 (2017).
[3] F. Flamini *et al.*, "Photonic architecture for reinforcement learning," arXiv: 1907.07503v1, 17 Jul. 2019.
[4] Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.
[5] A. Eldem *et al.*, "A model of deep neural network for Iris classification with different activation functions," 2018 International Conference on Artificial Intelligence and Data Processing (IDAP). DOI: 10.1109/IDAP.2018.8620866
[6] W. R. Clements *et al.*, "Optimal design for universal multiport interferometers," Optica **3**, 1460 (2016).
[7] G. Cong *et al.*, "High-efficient black-box calibration of large-scale Silicon photonics switches by bacterial foraging algorithm," OFC2019, M3B.3.
[8] G. Cong *et al.*, "Arbitrary reconfiguration of universal silicon photonic circuits by bacteria foraging algorithm to achieve reconfigurable photonic digital-to-analog conversion," Opt. Express **27**, 24914 (2019).
[9] K. Tanizawa *et al.*, "32×32 Strictly Non-Blocking Si-Wire Optical Switch on Ultra-Samll Die of 11×25 mm2," OFC2015, M2B.5.