Machine Learning Models for Predicting the Quality Factor of FSO Systems with Multiple Transceivers

Amal A. Algedir
School of Electrical and Computer Engineering
University of Oklahoma
Norman, United States
amal@ou.edu

Taissir Y. Elganimi
Electrical and Electronic Engineering Department
University of Tripoli
Tripoli, Libya
t.elganimi@uot.edu.ly

***Abstract*—Free space optical (FSO) communication is a promising solution to deliver the last mile communication and to guarantee a high data rate. However, the performance of FSO links can be significantly degraded by adverse weather conditions. Recently, machine learning algorithms (MLAs) have emerged for robust prediction to optimize the network performance. In this work, the Quality factor (Q) of FSO systems is estimated by means of four MLA models, namely, multi-linear regression, support vector regression, decision tree regression, and random forest regression. The synthetic data is used for training and testing these MLAs models, and several atmosphere conditions are considered with multiple transceivers FSO link system. The results of decision tree and random forest models demonstrated high coefficient of determination (R^2) and low mean square error (MSE) as compared to the other models.**

Keywords—Free Space Optics, machine learning, prediction, Q-factor, regression.

I. Introduction

Machine learning techniques for optical communications play a pivotal role that emerged recently and will continue as one of the intelligent techniques for the Sixth Generation (6G) and beyond wireless networks [1]. From the literature, it is shown that many of the state-of-the-art Machine Learning Algorithms (MLAs) have not yet been used in the field of optical communications. Thus, this research area is pristine. The prediction using MLAs has attracted the attention of researchers in different areas, such as education, communication, and industry. In this corresponding, a prediction algorithm is attempted to predict the system parameters by learning the pattern from the previous data. For example, in [2], MLAs are trained with past records of attendance of students to find a pattern of class attendance and to predict an accurate class strength.

Free Space Optics (FSO) communication technology is an emerging high speed point-to-point broadband technology. It is categorized under optical wireless communications and uses light as a transmission medium between the transceivers [3]. FSO systems can be deployed over a link distance of few kilometers. Over the last few years, FSO technology has gained a closer attention from researchers and acceptance in enterprise campus networks [3]. It provides high bandwidth with fast time and ease of installation as compared to fiber optics. Different weather conditions such as haze, fog and rain that cause suspended particles in the air are usually interacting with the photons of the optical wavelength, and subsequently causing a scattering of the optical signal [4]. The latter causes attenuation and reduces the overall availability of FSO systems. Additionally, FSO communication channel is affected by scintillation due to the effect of weather turbulence [5]. Moreover, diversity techniques are considered for improving the capacity performance of FSO channels under strong turbulence, and to compensate the effect of scintillation [6, 7].

Recently, FSO links with multiple transmitters and receivers architecture are becoming a more viable solution in order to improve the quality of FSO communication systems [8]. The Quality factor (Q-factor) in optical communications is an important parameter to evaluate the quality of the received optical signal [9]. Since the Q-factor of FSO communication systems is one of the main parameters that are used to determine the performance of FSO links, hence the effort here in this paper is to focus on the value of Q-factor of FSO links with various combinations of transmitters and receivers numbers.

The authors in [9] have analyzed the performance of FSO systems with single transceiver by predicting the Q-factor from the dataset with regression and classification. The model is validated with an experimental data, and it is shown that Support Vector Machine (SVM) regression model is much better than the other regression models to predict the Q-factor of FSO systems. In a recent work [10], unsupervised MLAs in FSO communication links are proposed in order to detect the number of concurrently transmitting users, where histogram and peak detection were employed in order to estimate the number of transmitting users sharing time slots and bandwidth simultaneously by exploiting their amplitude information.

This paper has utilized the supervised machine learning techniques to predict the Q-factor of FSO links. We considered multiple transceivers, the wavelength, and different weather conditions as input features to the MLAs models. Four MLAs, namely, Multi-linear Regression (MLR), Support Vector Regression (SVR), Decision Tree (DT) regression, and Random Forest (RF) regression are trained to predict the Q-factor. In addition, the accuracy of these four MLAs is computed in terms of the coefficient of determination (R^2) and the Mean Square Error (MSE).

The remainder of this paper is organized as follows. Section II describes the machine learning prediction models that are used in this paper. Section III presents the methodology which includes FSO system model and the simulation parameters. Section IV covers the simulation results. Finally, Section V gives the conclusion of the overall work.

978-1-7281-6264-5/20/$31.00 ©2020 IEEE

II. Machine Learning Prediction Models

Regression algorithms are applied in this paper in order to predict the Q-factor of FSO systems with multiple transceivers. The prediction model considers five independent variables (attributes) that represented by $\boldsymbol{X} = [x_1, x_2, x_3, x_4, x_5]$ as shown in Fig. 1. $\boldsymbol{X}$ is defined as the input attributes where each row is an instance to be applied to predict a continuous variable Q which is denoted by $(\hat{y}_i)$. For training each model, the K-fold cross validation is applied on the dataset $\{x_i; y_i\}$ to avoid the overfitting problem in some estimator models [11].

A. Multi-linear Regression (MLR)

MLR is an approach that models the relationship between independent variables and a dependent variable by fitting a linear equation. For the dataset of n statistical units, the linear regression model assumes that the relationship between the p-vector of regressors (x_i) and the dependent variable $(\hat{y})$ is linear. This relationship is modeled with a disturbance random error variable (e) that adds noise to the linear relationship. The best fitting is calculated by minimizing the sum of the squares errors. This model takes the following form [11]:

$$\hat{y} = B_0 + B_1 x_1 + B_2 x_2 + \cdots B_p x_p + e \qquad (1)$$

where B_0 denotes the value of y when all the independent variables, $x_1, x_2, \ldots, x_p$, are equal to zero, and $B_1, B_2, \ldots, B_\mathrm{p}$ are the estimated regression coefficients.

B. Support Vector Regression (SVR)

SVR is developed as a regression type of SVM. It is considered as a nonparametric technique because it relies on kernel functions. SVR model performs the regression task by finding the optimal regression hyperplane in which most of the training points lie within a margin (ε) around this hyperplane. Given $\{X; Y\}$, SVR model determines a function $f(x)$ that deviates from y_i by a value that is not greater than ε for each training point [12].

C. Decision Tree (DT) Regression

DT regression model is built as a tree structure. It is constructed from a root node using top-down induction, and involves splitting and breaking down the dataset into smaller subsets that contain similar samples (homogenous). The partitions of samples are performed by finding the attribute that returns the highest information gain. Entropy is used to calculate the homogeneity of a sample at each node of the tree. In addition, DT regression model performs very well when there is a non-linear and complex relationship between dependent and independent variables [11].

D. Random Forest (RF) Regression

RF regression algorithm is an ensemble learning technique that combines multiple regression trees on various sub-samples of the dataset. The output of this algorithm is the mean prediction (regression) of the individual trees. It uses averaging in order to improve the predictive accuracy, and to correct and control the over-fitting of the training set [13].

III. Methodology

In this paper, MLAs are utilized to estimate the Q-factor of FSO links under different weather conditions using synthetic data. The Q-factor provides the quality of the signal with respect to the distance of the signal from the noises. It covers the dispersions and nonlinearities. For the generation of the synthetic data, OptiSystem software is used in this paper in order to design the FSO system with multiple transceivers and to measure the Q-factor in different atmospheric disturbances. This software program is commonly used in optical communications, e.g. [9, 14].

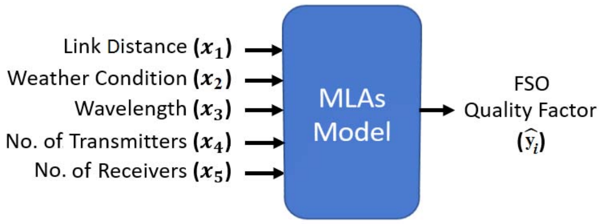

Fig. 1. Machine learning block diagram.

The simulation setup of FSO system model architecture with four transmitters and four receivers is shown in Fig. 2. This layout includes the optical transmitter unit that includes the Pseudo Random Bit Sequence (PRBS) generator, Non Return to Zero (NRZ) pulse generator, a Continuous Wave (CW) laser source, and Mach-Zehnder Modulator (MZM). In addition, the system model includes the FSO channels, optical amplifiers, Bit Error Rate (BER) analyzer, and optical power meter. The optical receiver contains a P-type Intrinsic N-type (PIN) photodiode with a Bessel Low Pass Filter (LPF). An LPF is used with a cut-off frequency of $(0.75 \times symbol\ rate)$. Moreover, a fork is used in this system to duplicate the input beam to all the channels. At the receiver side, the optical signals from all channels are combined with the help of a power combiner having multiple input ports. The simulation parameters including the atmospheric attenuation of each weather condition are listed in Table I.

IV. Simulation Results

MLAs are implemented in this paper using Python, and K-fold was used as a cross-validation. More specifically, a 10-

TABLE I. Simulation Parameters

Parameters	Values	
Atmospheric attenuation of each weather condition	Dense Fog	84.9 dB/Km
	Medium Fog	33.96 dB/Km
	Low Fog	15.55 dB/Km
	Dense Haze	4.285 dB/Km
	Medium Haze	1.54 dB/Km
	Low Haze	0.442 dB/Km
	Clear Air	0.2453 dB/Km
	Very Clear	0.0883 dB/Km
	Drizzle Rain	0.427 dB/Km
	Light Rain	1.982 dB/Km
	Average Rain	5.795 dB/Km
	Storm Rain	9.1996 dB/Km
	Strong Rain	23.182 dB/Km
Transmission rate	10 $Gbps$	
Transmitted power	20 dBm	
Wavelength	850 nm, 1310 nm and 1550 nm	
Number of transmitters	1, 2, 4 and 8	
Number of receivers	1, 2, 4 and 8	

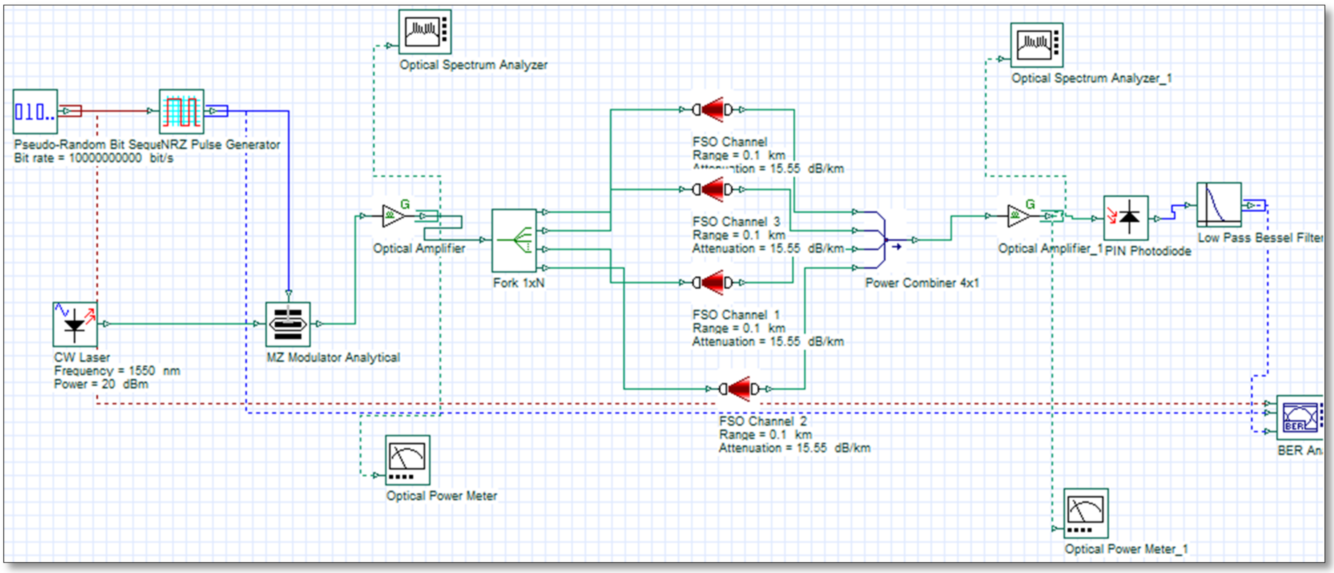

Fig. 2. FSO system model with four transmitters and four receivers.

fold split on a dataset of the size 2500×5 is used. Moreover, the model performance is assessed by two quantitative measurements, namely, the MSE of the estimated Q-factor and the coefficient of determination (R^2), which both are given as follows [3, 4]:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2}$$

$$R^2 = 1 - \frac{MSE}{SST} \tag{3}$$

where n is the number of samples, y_i is the true value of the variable, $\hat{y}_i$ is the predicted value of the variable that is predicted by ML model, and SST is the total sum of squares and given as [5]:

$$SST = \sum_{i=1}^{n}(y_i - \mu_y)^2 \tag{4}$$

where μ_y is the sample mean of the feature.

In Fig. 3, a comparison between MLAs models in terms of R^2 values is shown, while Fig. 4 illustrates a comparison of these models in terms of the MSE value. It is clear from these two figures that SVR with Gaussian kernel function shows the worst results where R^2 is equal to 0.09 and MSE is about 0.9. It is also clear that the MLR model does not capture all independent variables, and it cannot be used to estimate the Q-factor due to the fact that our data do not fulfill the pre-assumptions of the linearity between the dependent variable and the independent variables.

It can be also seen that both DT and RF regression models provided robust results in terms of R^2 coefficient and MSE as displayed in Figs. 3 and 4, respectively. The performance of RF with 10 estimators is also compared to DT and provided the best trade-off between performance and computational time. The lowest MSE value is achieved by RF regression model, and it was equal to 0.039 with R^2 of almost 0.953, while DT model provided an MSE value of about 0.092 and R^2 equals to 0.904. In contrast, the computational time of RF model was approximately 10 times more than that of DT model. Generally, RF model outperforms the DT model due to the fact that its performance is averaged using many DT models.

Since the number of the independent variables are more than three, the regression hyperplane cannot be visualized. Further insights of the regression models performance were gathered by plotting residuals. The residuals measure the difference between the actual value of the target variable (y) and the predicted value ($\hat{y}$). Figs. 5 and 6 illustrate the residuals of DT and RF models, respectively. It is clear from these two figures that the distribution of residuals does not seem to be completely random around the zero center line which confirms that linear regression cannot be applied. Alternatively, the DT and RF models have captured the non-linearity in the data by dividing the space into smaller sub-spaces. In addition, DT model can handle both categorical and numerical data. Furthermore, it is obvious that RF residual plot indicates a good improvement over the residual plot of DT model. Moreover, RF regression model has better generalization performance than DT regression due to the randomness that helps to reduce the variance of the model.

Fig. 7 shows the predicted and real values of the Q-factor using DT model for 50 data points. Generally, DT model is very accurate especially when the Q-factor is small. As the value of Q-factor increases, it can be clearly seen that a missprediction happens. Among the 50 data points, the worst missprediction was at about 0.22 and it is shown in the first data point.

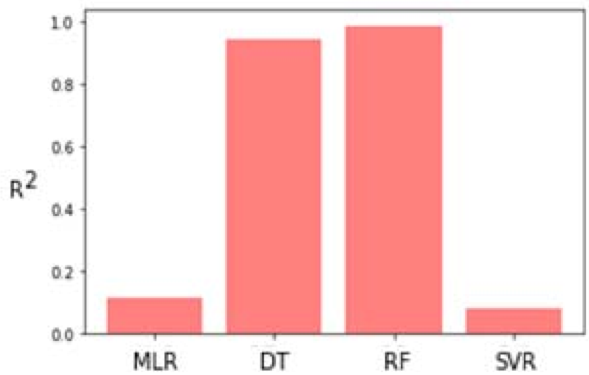

Fig. 3. Coefficient of determination of regression models.

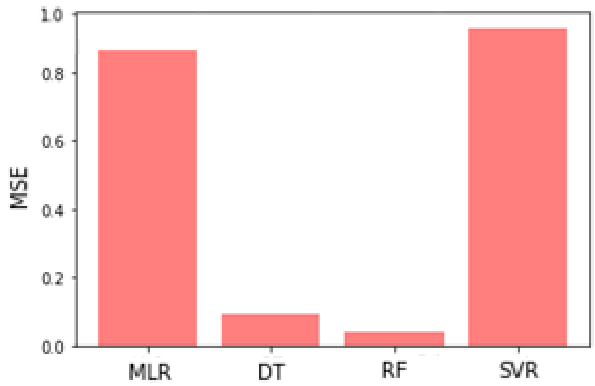

Fig. 4. Mean square error of regression models.

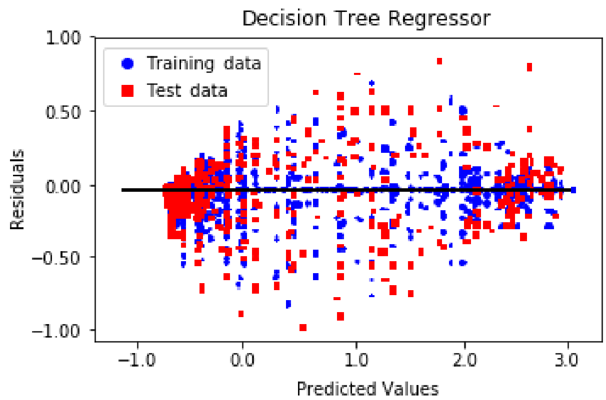

Fig. 5. Residual plot of DT regressor.

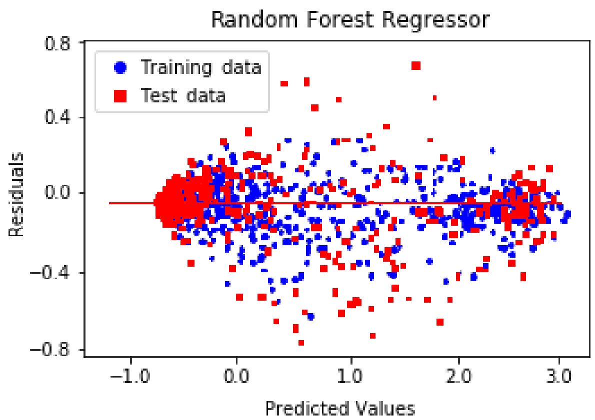

Fig. 6. Residual plot of RF regressor.

V. CONCLUSION

In this paper, different machine learning models are examined to predict the Q-factor of FSO links with multiple transceivers under different weather conditions. Based on the results, it is found that MLR and SVR models cannot be used to predict the Q-factor. Moreover, DT and RF models demonstrated an optimal estimation of the Q-factor. However, DT model would work better in real time, and it can be considered as a useful component to be integrated in decision tools of the network design to optimize the system performance.

ACKNOWLEDGMENT

The authors wish to thank Mrs. Rayan Ghouma for her valuable assistance in preparing the data used in this paper.

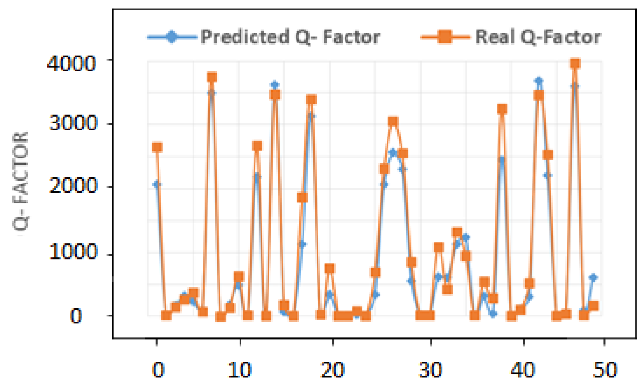

Fig. 7. Real value versus predicted value of Q-factor using DT model.

REFERENCES

[1] M. A. Amirabadi, “A survey on machine learning for optical communication [machine learning view],” Computer Science, Engineering, August 2019.

[2] E. Rashid, M. D. Ansari, V. K. Gunjan, and M. Khan, “Enhancement in teaching quality methodology by predicting attendance using machine learning technique,” Modern Approaches in Machine Learning and Cognitive Science: A Walkthrough, Springer, pp. 227-235, Feb. 2020.

[3] M. A. Khalighi and M. Uysal, “Survey on free space optical communication: A communication theory perspective,” IEEE Commun. Surveys Tuts., vol. 16, no. 4, pp. 2231-2258, 2014.

[4] F. Nadeem, V. Kvicera, M. S. Awan, E. Leitgeb, S. S. Muhammad, and G. Kandus, “Weather effects on hybrid FSO/RF communication link,” IEEE J. Sel. Areas in Commun., vol. 27, no.9, pp.1687-1697, Dec. 2009.

[5] H. Al Quwaiee, H.-C. Yang, and M.-S. Alouini. “On the asymptotic capacity of dual-aperture FSO systems with generalized pointing error model,” IEEE Trans. Wireless Commun., vol. 15, no. 9, pp. 6502-6512, Sept. 2016.

[6] A. H. Abd El-Malek, A. M. Salhab, S. A. Zummo, and M.-S. Alouini. “Security-reliability trade-off analysis for multiuser SIMO mixed RF/FSO relay networks with opportunistic user scheduling,” IEEE Trans. Wireless Commun., vol. 15, no. 9, pp.5904-5918, Sept. 2016.

[7] L. Yang, X. Gao, and M.-S. Alouini. “Performance analysis of free-space optical communication systems with multiuser diversity over atmospheric turbulence channels,” IEEE Photonics Journal, vol. 6, no. 2, pp.1-17, April 2014.

[8] N. H. M. Noor, W. Al-Khateeb, and A.W. Naji, “Experimental evaluation of multiple transmitters/receivers on free space optics link,” In 2011 IEEE Student Conference on Research and Development (SCOReD), pp. 128-133, Cyberjaya, Malaysia, 19-20 Dec. 2011.

[9] K. F. Rahman, F. Afrin, and F. S. bin Taip, “Optical network performance analysis through prediction of Q-factor on simulated and validated dataset,” 2018 IEEE International Conference on Artificial Intelligence in Engineering and Technology, Kota Kinabalu, Malaysia, 8 Nov. 2018.

[10] F. Aveta, H. H. Refai, and P. G. Lopresti, “Number of users detection in multi-point FSOC using unsupervised machine learning,” IEEE Photonics Techn. Letters, vol. 31, no. 22, pp. 1811-1814, Nov. 2019.

[11] E. Alpaydin, “Introduction to Machine Learning,” 3rd ed., 2014.

[12] C. Cortes and V. Vapnik, “Support-vector networks,” Machine Learning, vol. 20, no. 3, pp. 273-297, Nov. 1995.

[13] Z. Liu, and H. Chen, “A predictive performance comparison of machine learning models for judicial cases,” 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Honolulu, HI, USA, 27 Nov.-1 Dec. 2017.

[14] F. Hossain and Z. Afroze, “Eliminating the effect of fog attenuation on FSO link by multiple TX/RX system with travelling wave semiconductor optical amplifier,” in 2013 2nd IEEE International Conference on Advances in Electrical Engineering (ICAEE), Dhaka, Bangladesh, 19-21 Dec. 2013.