Shareef, A. A., Kaur, J., Imran, M. A., Ali, H. T. M., Abbasi, Q. H. and Abbas, H. T. (2023) A Statistical Analysis of Feature Transformation for Efficient Localisation in Urban Environments. In: 2023 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting, Portland, Oregon, USA, 23–28 July 2023, pp. 269-270. ISBN 9781665442282.

A Statistical Analysis of Feature Transformation for Efficient Localisation in Urban Environments

Azad Adil Shareef(1), Jaspreet Kaur(2), Muhammad A Imran(2), Haithem Taha Mohammed Ali(3), Qammer H. Abbasi(2), and Hasan T Abbas(2)

(1) University of Duhok, Kurdistan, Iraq. Email; azada@uod.ac
(2) James Watt School of Engineering, University of Glasgow, Glasgow, United Kingdom, G12 8QQ
(3) University of Zakho, Nawroz University, Kurdistan, Iraq.

***Abstract*—In this paper, we perform a transformation-based statistical analysis with an eye to designing a robust and efficient localisation scheme. To this end, we evaluate the coefficient of determination (COD) also denoted as R^2 on simulated electromagnetic wave propagation models in an urban environment. Our transformation-based statistical models show that two measurable network parameters, namely the received power (RP) and the time of arrival (ToA), present a strong correlation with a mobile user's given location. By transforming the network parameters we were able to achieve COD of 0.577 for the RP and 0.549 for ToA using the Modulus transformation. We believe that by exploiting the high correlation of parameter transformation, there is potential to design fast and robust machine learning localisation schemes through which the future location of a mobile user can be accurately and reliably predicted.**

I. Introduction

In the context of a wireless network, localisation is the process of finding the location of user equipment (UE) based on the available network parameters. Artificial intelligence (AI) is forecast to be an essential component of future-generation wireless communication systems. There have been several studies in which the performance of various modalities of a communication system was dramatically improved using AI. In [1], the authors proposed an intelligent transportation system that predicts angular and temporal channel states using out-of-spectrum. In [2], Alkhateeb applied deep learning on a simulated dataset to solve hard network propagation optimisation models. Similarly, Udita et al [3] developed a deep learning prediction scheme in which network parameters such as received power (RP) and time of arrival (ToA) were utilised to develop an accurate location estimation scheme. Although the performance of AI-based location prediction schemes is remarkably high, the practical deployment of such a scheme is still far from reality owing to the requirements of a huge dataset for training purposes, as well as the high computational cost of learning. Moreover, applications such as V2X networks demand real-time and on-the-spot decision-making even under previously unseen conditions.

Feature engineering involves preprocessing procedures such as dimensionality reduction, and therefore, the removal of irrelevant and redundant data is a highly effective process to increase the performance of machine learning algorithms [4]. Furthermore, it is notable that most studies that develop a neural network-based prediction model resort to a regression approach where the goal is to attain a coefficient of determination (CoD) or R^2 value as close to 1. In this paper, we quantify the predictive power of wireless network parameters using transformation-based statistical models. A higher R^2 as a result of transformation can lead to a more accurate AI model that is simultaneously computationally efficient. The network data is collected using a ray-tracing simulation performed in an urban environment.

II. Transformation Models

A. Box-Cox Transformation

The first to set the mathematical approach to the power transformations was Box and Cox [5]. Power transformations are a collection of algorithms that apply a power function (such as a logarithm or exponent) to make a variable's probability distribution Gaussian or more Gaussian-like. Normality is an important assumption for many statistical models. A one-parameter (λ) Box-Cox transformation (BCT) on a parameter y is expressed as,

$$\mathrm{BCT}(y) = \begin{cases} y^{\lambda-1}/\lambda & \text{if } \lambda \neq 0 \\ \log y & \text{if } \lambda = 0 \end{cases} \tag{1}$$

B. Exponential data transformation

A unique feature of Exponential data transformation (EDT) is that it permits negative parameter values. The main motivation to use this kind of transformation is to enhance the linearity of the data. The EDT parameter family is identified as [6],

$$\mathrm{EDT}(y) = \begin{cases} [\exp(y \times \lambda) - 1]/\lambda & \text{if } \lambda \neq 0 \\ y & \text{if } \lambda = 0 \end{cases} \tag{2}$$

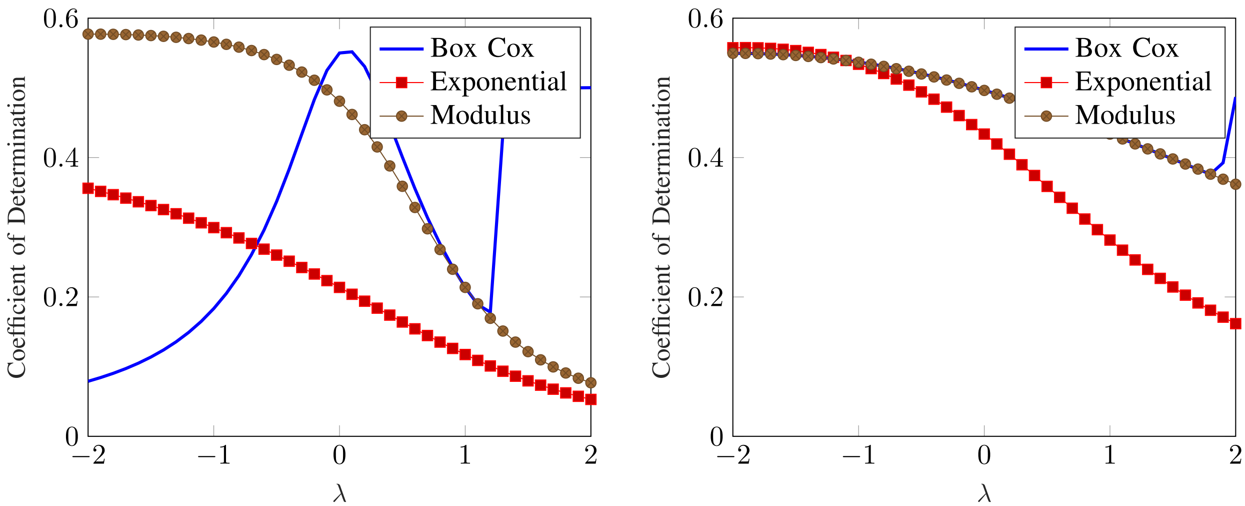
C. Modulus Transformation

The Modulus transformation (MT) is a generalised form of the BCT as both positive and negative data values can be used. Moreover, it removes skewness from the data

$$\mathrm{MT}(y) = \begin{cases} \mathrm{sign}\left\{\left(\mathrm{abs}(y)+1\right)^{\lambda} - 1\right) / \lambda\right\} & \text{if } \lambda \neq 0 \\ \mathrm{sign}\{\log(abs(y)+1\} & \text{if } \lambda = 0 \end{cases} \tag{3}$$

III. Results and Discussion

We compare the three different transformations; BCT, EDT, and MT to estimate power parameters using COD. A simulated dataset was generated using Wireless Insite 3.3 on a personal computer where the University of Glasgow Gilmorehill Campus was chosen as the urban environment. Further details of



(a) Received Power

(b) Time of Arrival

Fig. 1: R^2 curves of RP and ToA for the three transformations used.

the generated dataset can be found in [7]. We used RStudio 4.0.5 2022 software and SPSS version 25 to develop the statistical models. The $\mathrm{Path-ID}$ that denotes the position of a UE in the dataset was chosen as the independent variable. We used COD as a criterion for choosing the best value of the transformation parameter. COD is a squared multiple correlation coefficient which is a useful index in regression analyses. Some scientists use the coefficient of determination as a measure of goodness of fit and precision in prediction for the general linear model [8]. The COD for the RP as the dependent variable and $\mathrm{Path-ID}$ as an independent variable for the original wireless data set is 0.213. Furthermore, the COD for the ToA as the dependent variable and $\mathrm{Path-ID}$ as an independent variable for the original wireless data set is 0.433. Table I shows the optimal values of the power parameter, λ which corresponds to different transformations by using COD. This study shows that the optimal value of COD for RP when λ = 0.1, -2, and -2 is 0.551, 0.355, and 0.577 by using BCT, EDT, and MT, respectively in Fig. 1. In addition, the optimal value of COD for the TOA when λ = -2,-1.9 and 2 is 0.549, 0.557, and 0.549 by using BCT, EDT, and MT, respectively. Although for classification purposes, the desired COD should be ideally as close to 1, however, the dataset contains non-line of sight parameters, the prediction of which is particularly hard as has been reported elsewhere [3], [7].

TABLE I: The result of optimal value of power paramete

Transformations	Optimal λ	
	Received Power	Time of Arrival
BCT	0.1	-2
EDT	-2	-1.9
MT	-2	-2

IV. Conclusion

The authors believe that it is an improvement of an optimal value that meets the conditions of the COD for estimating linear regression model of the original response vector. We can conclude that MT that correspond to λ = -2 is an optimal value of the power parameter using the COD criteria to apply it on wireless dataset for both models RP and TOA. We will use the transformed data in our localisation algorithm for improving location estimation accuracy.

Acknowledgement

AAZ acknowledges the financial support provided by the University of Duhok, for a PhD researcher exchange to the University of Glasgow.

References

[1] C. K. Anjinappa and I. Guvenc, "Angular and Temporal Correlation of V2X Channels across Sub-6 GHz and mmWave Bands," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, May 2018, pp. 1–6.

[2] A. Alkhateeb, "DeepMIMO: A Generic Deep Learning Dataset for Millimeter Wave and Massive MIMO Applications," Feb. 2019.

[3] U. Bhattacherjee, C. K. Anjinappa, L. Smith, E. Ozturk, and I. Guvenc, "Localization with Deep Neural Networks using mmWave Ray Tracing Simulations," Feb. 2020.

[4] S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," in *2014 Science and Information Conference*, Aug. 2014, pp. 372–378.

[5] G. E. P. Box and D. R. Cox, "An Analysis of Transformations," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 26, no. 2, pp. 211–252, 1964.

[6] J. A. John and N. R. Draper, "An Alternative Family of Transformations," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 29, no. 2, pp. 190–197, 1980.

[7] J. Kaur, O. R. Popoola, M. A. Imran, Q. H. Abbasi, and H. T. Abbas, "Deep Neural Network for Localization of Mobile Users using Raytracing," in *2022 IEEE USNC-URSI Radio Science Meeting (Joint with AP-S Symposium)*. Denver, CO, USA: IEEE, Jul. 2022, pp. 76–77.

[8] J. P. Barrett, "The Coefficient of Determination—Some Limitations," *The American Statistician*, vol. 28, no. 1, pp. 19–20, Feb. 1974.