ISSN 2347 - 3983

Volume 8. No. 7, July 2020

International Journal of Emerging Trends in Engineering Research

Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter37872020.pdf

https://doi.org/10.30534/ijeter/2020/37872020

Deep Divergence-Based Clustering of Wireless Multipaths for Simultaneously Addressing the Grouping and the Cardinality

Jojo Blanza[1,2], Lawrence Materum[1,3], Takuichi Hirano[3]

[1]Department of Electronics and Communications Engineering, De La Salle University, 2401 Taft Ave., Malate, Manila, 0922, Philippines, jojo_f_blanza@dlsu.edu.ph

[2]Electronics Engineering Department, University of Santo Tomas, Espana Blvd., Sampaloc, Manila, 1008, Philippines, jfblanza@ust.edu.ph

[3]Tokyo City University, 1-28-1 Tamazutsumi, Setagaya, Tokyo, 158-8557, Japan

ABSTRACT

Deep divergence-based clustering (DDC) is used to cluster COST 2100 channel model (C2CM) wireless propagation multipaths. The dataset is taken from the IEEE DataPort. DDC solves the membership of the clusters. DDC builds on information theoretic divergence measures and geometric regularization in order to determine the membership of the clusters. The cluster count is then computed through the cluster-wise Jaccard index of the membership of the multipaths to their clusters. The performance of DDC is evaluated using the Jaccard index by comparing the reference multipathdatasets from IEEE DataPort with the calculated multipath clusters obtained by DDC. Results show that DDC can be used as an alternative clustering approach in the field of channel modeling.

Keywords: channel models, clustering methods, data handling, data models, data preprocessing, multipath channels, radiowave propagation

1. INTRODUCTION

The European Cooperation in Science and Technology (COST) 2100 Channel Model (C2CM) [1] can replicate the stochastic properties of multiple-input multiple-output (MIMO) wireless propagation channels. C2CM is characterized by multipath clusters. Groups of multipath components with similar delay and angles comprise a multipath cluster. A multipath component (MPC) is classified based on the delay, angle of departure (Azimuth of Departure (AoD), Elevation of Departure (EoD)), and angle of arrival (Azimuth of Arrival (AoA), Elevation of Arrival (EoA)). Channel modeling is used to analyze the characteristics of wireless communications systems. The attribute, performance, and efficiency of the communications system can be studied, understood, and improved by the generated channel model. Several channel models and measurements reveal the clustering of multipaths. The accuracy and correctness of the channel models significantly affect the precision and exactness of clustering wireless propagation multipaths.

Clustering aims to find the underlying structure of the data and group similar data together. It finds application in a wide range of domains such as, but not limited to, intrusion detection [2] and feature selection [3]. Numerous clustering approaches [4]–[13] have been used to find the number of clusters in multipaths. However, they have not considered the membership of the clusters. The problem of providing only just the number of clusters is that there is no assurance of the correctness of the membership of the multipath clusters even though the number of clusters is accurate. Inaccurate clustering of the wireless propagation multipaths leads to incorrect channel models and thereby degradation in performance. In this work, the potential of DDC is illustrated in order to address this issue by clustering the wireless propagation multipaths.

Analyzing wireless multipaths is an important problem where clustering is crucial. Previous work has been limited to shallow approaches [12], while deep neural networks have found application for supervised wireless multipath processing [14]–[15]. At the same time, there has been a recent line of work that has tried to investigate the use of neural networks for clustering [12]. In this work, to exploit the ability of deep neural networks to learn feature representations, we explore the potential of deep clustering for the task of grouping wireless multipaths.

This study shows the results of DDC in clustering C2CM datasets taken from IEEE DataPort. The work presents the clustering of wireless propagation multipaths using DDC by solving the clustering accuracy based on the Jaccard index of the membership of the clusters, and concurrently the cluster count based on a proper initial estimate. The main contributions of this study are (1) DDC, which provide the membership of the clusters, is applied for the first time to cluster C2CM wireless propagation multipaths taken from the IEEE DataPort; and (2) the results show that the clustering approach can be used as an alternative in studying channel models.

The paper is organized in the following way. Section 2 describes the datasets used in clustering multipaths. Section 3 discusses DDC as the multipath clustering approach. Section

4 presents the results in clustering multipaths obtained by DDC. Section 5 concludes the study.

2. COST 2100 MULTIPATH CLUSTER DATASETS

A channel impulse response that is changing with time (designated by t) is the group of MPCs from all the multipath clusters according to the location of the mobile station (MS) in relation to the base station (BS). The channel impulse response is based on the delay and direction domain and is given as

$$h(t,\tau,\mathbf{Y}^{\mathrm{BS}},\mathbf{Y}^{\mathrm{MS}}) = \sum_{k=1}^{K}\sum_{p} a_{k,p}\delta(\tau-\tau_{k,p})\delta(\mathbf{Y}^{\mathrm{BS}}-\mathbf{Y}_{k,p}^{\mathrm{BS}})\delta(\mathbf{Y}^{\mathrm{MS}}-\mathbf{Y}_{k,p}^{\mathrm{MS}}) \quad (1)$$

where $a_{k,p}$ is the complex amplitude of the pth MPC in the kth cluster, K is the set of visible cluster indexes, $\mathbf{Y}_{k,p}^{MS}$ is the direction of arrival (AoA, EoA) of the MPC, and $\mathbf{Y}_{k,p}^{BS}$ is the direction of departure (AoD, EoD). The complete discussion of C2CM and the detailed description of multipath components and multipath clusters can be found in [1], while [16] summarizes the study for easier understanding.

The indoor and semi-urban scenario wireless channel propagation multipaths are obtained from the multipath cluster datasets in IEEE DataPort [17]. The common dataset can be applied in modeling wireless channels, assessing clustering approaches, and validating clustering accuracies. The detailed discussion on the generation of the datasets can be found in [18]. The following are the reference data used in clustering multipaths:

1. Indoor, band 1 (B1), line-of-sight (LOS), single link (SL)
2. Indoor, band 2 (B2), line-of-sight, single link
3. Semi-Urban, band 1, line-of-sight, single link
4. Semi-Urban, band 2, line-of-sight, single link
5. Semi-Urban, band 1, line-of-sight, multiple links (ML)
6. Semi-Urban, band 2, line-of-sight, multiple links
7. Semi-Urban, band 1, non line-of-sight, single link
8. Semi-Urban, band 2, non line-of-sight, single link

There are thirty trials for each channel scenario. Each trial has a different number of multipaths and multipath clusters. The power component is eliminated since it is not required in the process of clustering. The cluster identifications or IDs are also removed in the clustering process as they are only utilized as reference IDs in evaluating the computed IDs.

The Jaccard index which serves as the similarity measure is calculated as

$$\eta = \frac{|\boldsymbol{C}_{\mathrm{ref}} \cap \boldsymbol{C}_{\mathrm{calc}}|}{|\boldsymbol{C}_{\mathrm{ref}} \cup \boldsymbol{C}_{\mathrm{calc}}|} = \frac{M_{11}}{M_{11}+M_{10}+M_{01}} \quad \in [0,1] \quad (2)$$

where $|\cdot|$ refers to cardinality, $\boldsymbol{C}_k \in \boldsymbol{C}$, $K = |\boldsymbol{C}|$ is the number of multipath clusters, $\boldsymbol{C}_{\mathrm{ref}}$ is the reference clusters, and $\boldsymbol{C}_{\mathrm{calc}}$ is the calculated clusters. M_{11} is the total number of multipath clusters for the accuracy on the number of clusters or the total number of multipaths for the accuracy on the membership of the clusters in $\boldsymbol{C}_{\mathrm{ref}}$ that is the same as in $\boldsymbol{C}_{\mathrm{calc}}$. On the other hand, M_{10} is the total number of multipath clusters for the accuracy on the number of clusters or the total number of multipaths for the accuracy on the membership of the clusters in $\boldsymbol{C}_{\mathrm{ref}}$ that are not in $\boldsymbol{C}_{\mathrm{calc}}$. Lastly, M_{01} is the total number of multipath clusters for the accuracy on the number of clusters or the total number of multipaths for the accuracy on the membership of the clusters in $\boldsymbol{C}_{\mathrm{calc}}$ that are not in $\boldsymbol{C}_{\mathrm{ref}}$. The Jaccard index ranges between 0 and 1, with one being the highest. A Jaccard index of 1 means that the reference multipath clusters are the same as the calculated multipath clusters, or the membership of the reference multipath clusters is the same as the membership of the calculated multipath clusters. On the other hand, a Jaccard index of 0 means that there is no calculated multipath cluster that is the same as the reference multipath clusters, or the membership of the calculated multipath clusters is not equal to the membership of the reference multipath clusters.

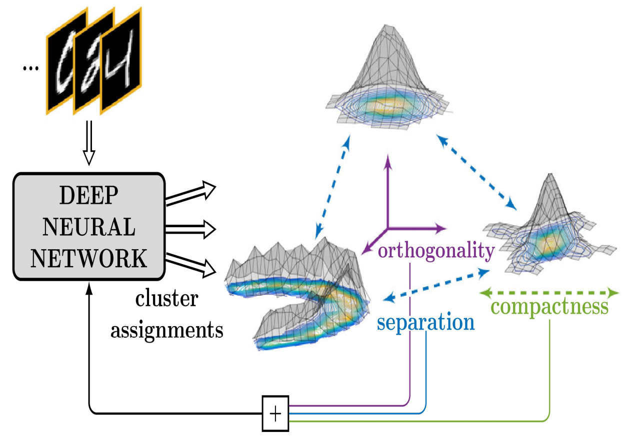

Figure 1: Fundamental objectives of divergence as a dissimilarity measure [19]

3. DEEP DIVERGENCE-BASED CLUSTERING (DDC)

DDC [19] optimizes a loss function defined by

$$L = \frac{1}{k}\sum_{i=1}^{k-1}\sum_{j>i}\frac{\boldsymbol{\alpha}_i^T\boldsymbol{K}_{\mathrm{hid}}\boldsymbol{\alpha}_j}{\sqrt{\boldsymbol{\alpha}_i^T\boldsymbol{K}_{\mathrm{hid}}\boldsymbol{\alpha}_i\boldsymbol{\alpha}_j^T\boldsymbol{K}_{\mathrm{hid}}\boldsymbol{\alpha}_j}} + \mathrm{triu}(\boldsymbol{A}\boldsymbol{A}^T) + \frac{1}{k}\sum_{i=1}^{k-1}\sum_{j>i}\frac{\boldsymbol{m}_i^T\boldsymbol{K}_{\mathrm{hid}}\boldsymbol{m}_j}{\sqrt{\boldsymbol{m}_i^T\boldsymbol{K}_{\mathrm{hid}}\boldsymbol{m}_i\boldsymbol{m}_j^T\boldsymbol{K}_{\mathrm{hid}}\boldsymbol{m}_j}} \quad (3)$$

where k is the number of clusters, $\boldsymbol{K}_{\mathrm{hid}}$ is the kernel similarity matrix, $\boldsymbol{\alpha}$ is the soft cluster assignment, $\boldsymbol{A}$ is the cluster assignment matrix, $\mathrm{triu}(\boldsymbol{A}\boldsymbol{A}^T)$ is the upper triangular of $\boldsymbol{A}\boldsymbol{A}^T$ and $\boldsymbol{m}$ is the simplex corner assignment.

Figure 1 shows divergence as a dissimilarity measure. The divergence measure achieves compactness within clusters and separation between clusters. The orthogonality, on the other hand, comes from the regularization term $\mathrm{triu}(\boldsymbol{A}\boldsymbol{A}^T)$.

The schematic diagram in clustering COST 2100 multipaths using DDC is shown in Figure 2. The algorithm supports

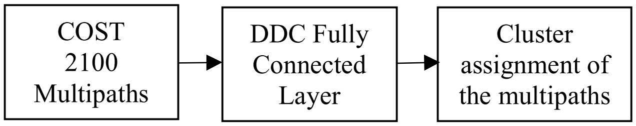

Figure 2: Schematic depiction of DDC in clustering multipaths

Table 1: Jaccard indices for the membership of clusters and cluster count

Channel Scenario	Membership of Clusters	Cluster Count
Indoor Band 1	0.7898	0.7390
Indoor Band 2	0.8210	0.7607
Semi-Urban Band 1 Line-of-Sight Single Link	0.2488	0.0121
Semi-Urban Band 2 Line-of-Sight Single Link	0.2433	0.0181
Semi-Urban Band 1 Non Line-of-Sight Single Link	0.2464	0.0055
Semi-Urban Band 2 Non Line-of-Sight Single Link	0.2381	0.0077
Semi-Urban Band 1 Line-of-Sight Multiple Links	0.1637	0.0108
Semi-Urban Band 2 Line-of-Sight Multiple Links	0.1631	0.0070

end-to-end learning, does not need a pre-training phase, and explicitly exploits the geometry of the output space to solve the optimization problem. The wireless propagation multipaths generated by C2CM serve as the input to the clustering approach. The input data consisting of the x-, y-, z-components of AoD, x-, y-, z-components of AoA, and delay with a different number of multipaths and multipath clusters are grouped by DDC. The number of classes entered in DDC is based on the cluster-ID assignments generated by C2CM. This input is crucial since the accuracy of DDC for multipath clustering the C2CM datasets can then be measured. The membership of clusters is solved by DDC. The cluster count is then calculated according to the membership of the multipaths to their clusters.

4. RESULTS OF CLUSTERING MULTIPATHS

The Jaccard indices for both the membership of clusters and the cluster count of the eight-channel scenarios are shown in Table 1. A smaller number of multipaths and multipath clusters are generated by the indoor channel scenarios due to limited reflections of signals by the enclosed space. For this reason, the indoor channel scenarios have better accuracy for both the membership of clusters and the cluster count compared with the semi-urban channel scenarios. On the other hand, the semi-urban channel scenarios generated a higher number of multipaths and multipath clusters by the wider surroundings due to more number of interacting objects that reflect the signals. Consequently, lower accuracy for both the membership of clusters and the cluster count is attained in semi-urban channel scenarios.

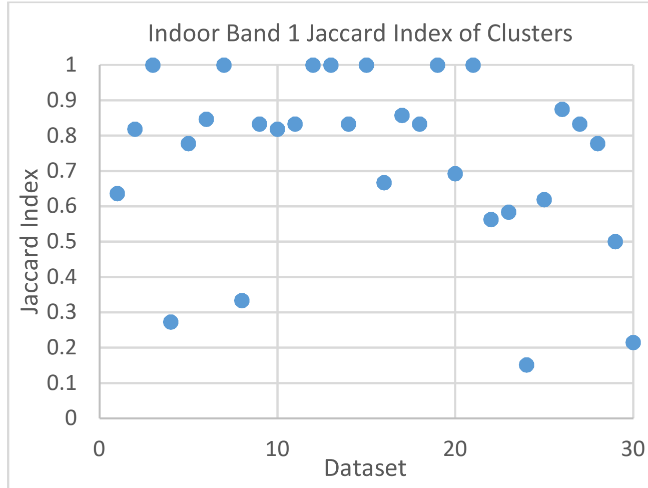

Figure 3: Jaccard index of the cluster count in indoor band 1 line-of-sight single link

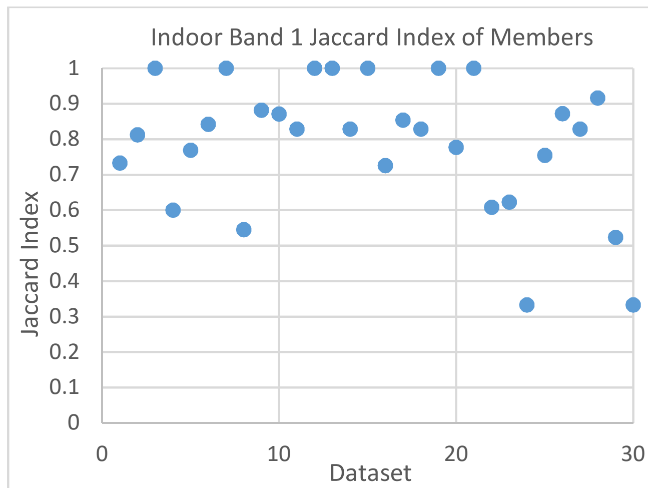

Figure 4: Jaccard index of membership of clusters in indoor band 1 line-of-sight single link

Figure 3 illustrates the Jaccard index of the cluster count in indoor band 1 line-of-sight single link. The minimum index is 0.1515, the maximum index is 1, and the mean index is 0.7390. Figure 4 shows the Jaccard index of the membership of the clusters of indoor band 1 line-of-sight single link. The minimum index is 0.3333, while the maximum index is 1. The mean index is 0.7898. Figure 5 presents the Jaccard index of the cluster count in indoor band 2 line-of-sight single link. The minimum index is 0.2, the maximum index is 1, and the mean index is 0.7607. Figure 6 displays the Jaccard index of the membership of the clusters of indoor band 2 line-of-sight single link. The minimum index is 0.4217, while the maximum index is 1. The mean index is 0.8210. Figure 7 demonstrates the Jaccard index of the cluster count of semi-urban band 2 line-of-sight single link. The minimum index is 0.2, the maximum index is 1, and the mean index is 0.0181. Figure 8 reveals the Jaccard index of the membership of clusters of semi-urban band 2 line-of-sight single link. The minimum index is 0.4217, while the maximum index is 1. The mean index is 0.2433. Figure 9 indicates the Jaccard index of the cluster count of semi-urban band 1 line-of-sight multiple

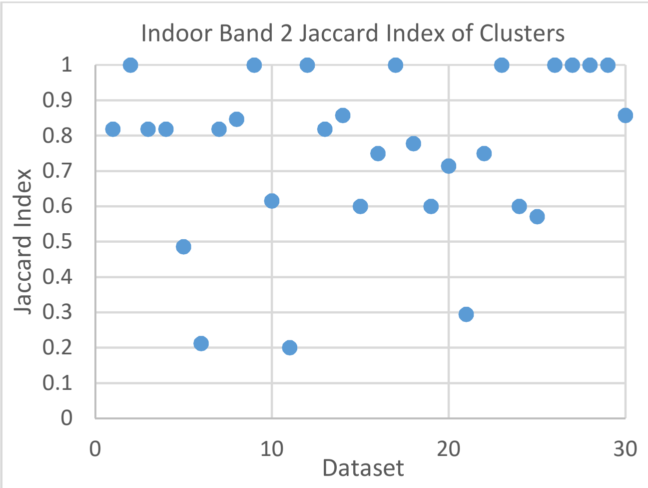

Figure 5: Jaccard index of the cluster count in indoor band 1 line-of-sight single link

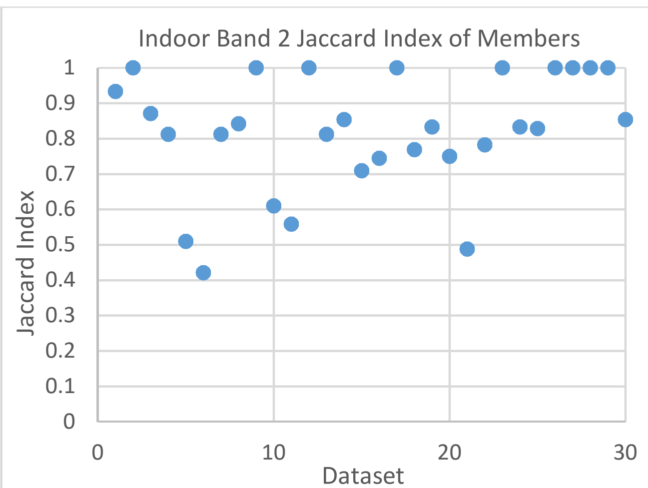

Figure 6: Jaccard index of membership of clusters in indoor band 1 line-of-sight single link

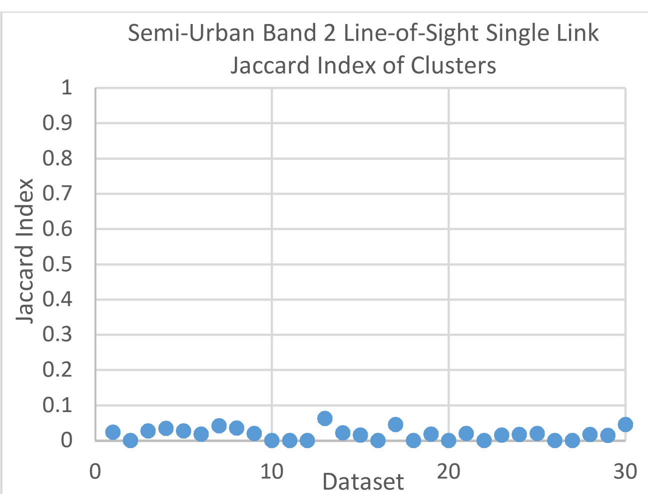

Figure 7: Jaccard index of the cluster count in semi-urban band 2line-of-sight single link

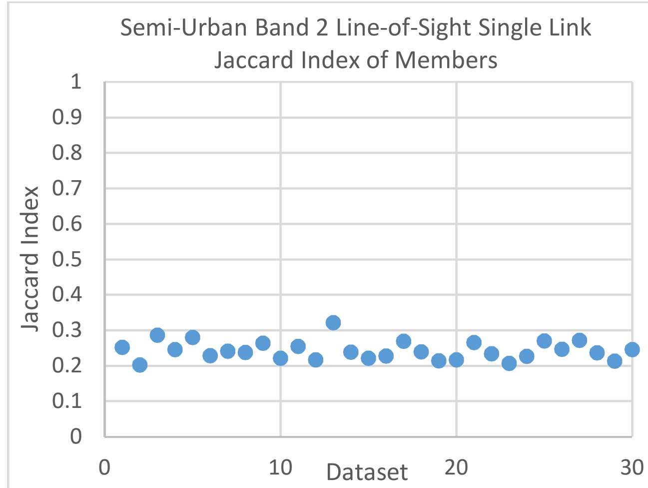

Figure 8: Jaccard index of membership of clusters in semi-urban band 2 line-of-sight single link

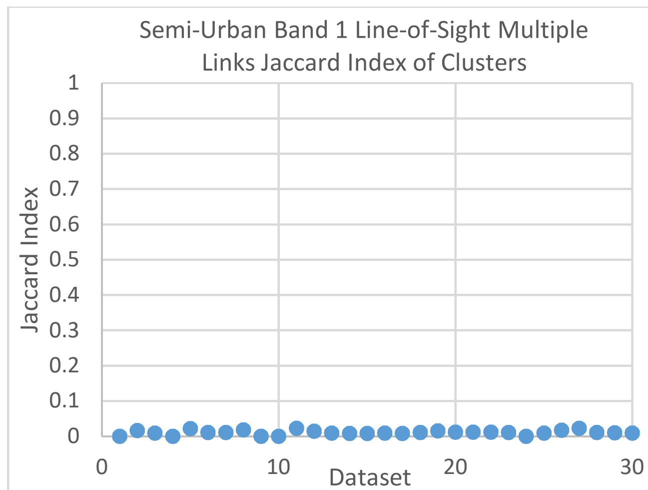

Figure 9: Jaccard index of the cluster count in semi-urban band 1 line-of-sight multiple links

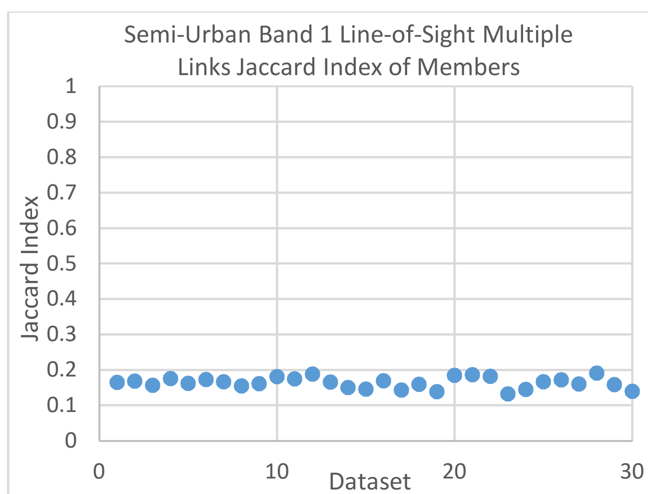

Figure 10: Jaccard index of membership of clusters in semi-urban band 1 line-of-sight multiple links

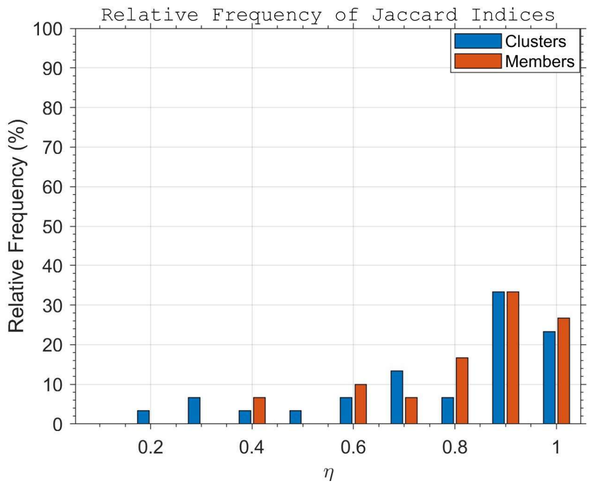

Figure 11: Relative frequency of Jaccard indices for indoor band 1 line-of-sight single link

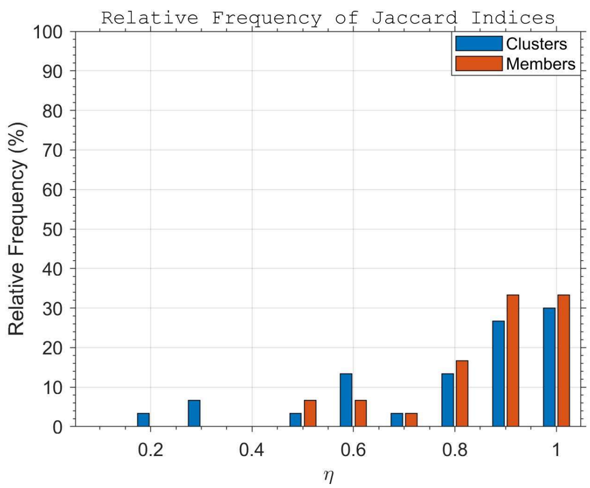

Figure 12: Relative frequency of Jaccard indices for indoor band 2 line-of-sight single link

links. The minimum index is 0, the maximum index is 0.0227, and the mean index is 0.0108. Figure 10 exhibits the Jaccard index of the membership of clusters of semi-urban band 1 line-of-sight multiple links. The minimum index is 0.1321, while the maximum index is 0.1906. The mean index is 0.1637.

Figure 11 illustrates the relative frequency of the Jaccard indices on the cluster count and the membership of the clusters for indoor band 1 line-of-sight single link. For the cluster count (in blue), 56.66% of the indices occurred from 0.8 to 1, while 60% of the index for the membership of the clusters (in red). Figure 12 shows the relative frequency of the Jaccard indices on the cluster count and the membership of the clusters for indoor band 2 line-of-sight single link. For the cluster count (in blue), 56.67% of the indices are greater than 0.8, while 66.66% for the membership of the clusters (in red). Figure 13 presents the relative frequency of the Jaccard indices for semi-urban band 2 line-of-sight single link. For the cluster count (in blue), all of the indices are less than 0.1, while for the membership of the clusters (in red) all indices

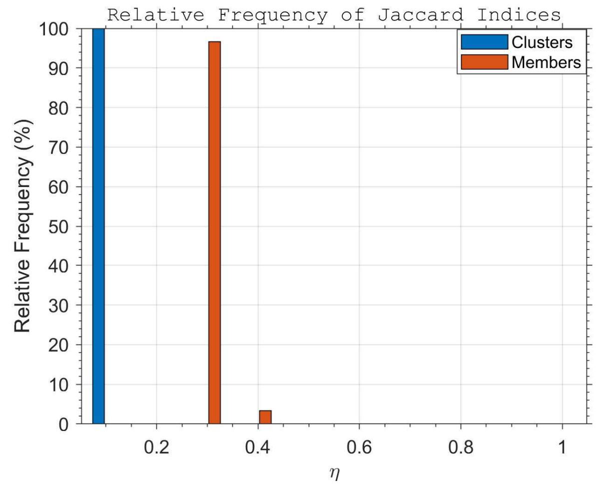

Figure 13: Relative frequency of Jaccard indices for semi-urban band 2 line-of-sight single link

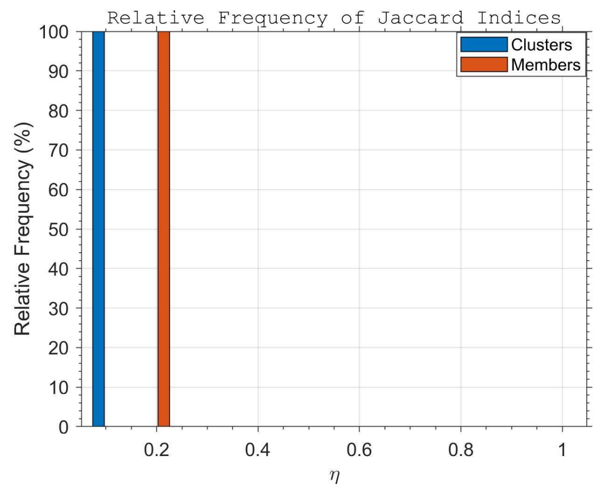

Figure 14: Relative frequency of Jaccard indices for semi-urban band 1 line-of-sight multiple links

are within the 0.2 to 0.4 range. Figure 14 displays the relative frequency of the Jaccard indices for semi-urban band 1 line-of-sight multiple links. All of the indices fall within the 0 to 0.1 range for the cluster count (in blue), while all of the indices of the membership of clusters (in red) are between 0.2 and 0.3.

The results show that DDC can determine the membership of the clusters. DDCdoes not directly determine the cluster count but through the cluster-wise Jaccard index of the membership to their clusters. A priori information on the cluster count is unknown, but once known, concurrent determination of the membership and cluster count becomes possible.

DDC clusters the indoor multipaths decently when fewer multipaths and multipath clusters characterize the channel model. At the same time, DDC doesn't fare well in semi-urban scenarios when there are more multipaths and multipath clusters present. Improvements can be made for DDC to perform well and increase the accuracy in clustering

C2CM multipaths, especially in the semi-urban scenarios. Overall, DDC can be applied in the field of channel modeling as an alternative multipath clustering approach.

5.CONCLUSION

This work presents the results of DDC in clustering C2CM wireless propagation multipaths. It is the first time that DDC is applied in multipath clustering. DDC solves the membership of the clusters based on the Cauchy-Schwarz divergence as the measure of dissimilarity between clusters, and the cluster count is computed based on the membership of multipaths to their clusters. The data obtained from IEEE DataPort are used as reference data for the C2CM multipath clusters. The calculated clusters obtained by DDC are compared with the reference clusters using the Jaccard index to evaluate the performance clustering accuracy of DDC. Results show that DDC clusters well the multipaths in indoor scenarios but performs poorly in semi-urban scenarios. Thus, DDC can be used in the field of channel modeling as an alternative clustering approach, but improvements are necessary to enhance multipath clustering accuracy. Other options can be considered based on the results of clustering multipaths with solving simultaneously the cluster count and the membership clusters taken into account.

ACKNOWLEDGEMENT

The authors would like to thank the Department of Science and Technology-Engineering Research and Development for Technology (DOST-ERDT), Department of Science and Technology-Advanced Science and Technology Institute-Computing and Archiving Research Environment (DOST-ASTI-COARE), and De La Salle University. Dr. Michael Kampffmeyer is acknowledged for his valuable comments.

REFERENCES

1. R. Verdone and A. Zanella.**Pervasive Mobile and Ambient Wireless Communications: COST Action 2100**, Springer Science & Business Media, 2012, ch. 3.
2. G. Pavithra, P. Abirami, S. Bhuvaneshwari, S. Dharani, and B. Haridharani. **A Survey on Intrusion Detection Mechanism using Machine Learning Algorithms**, *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 4, pp. 945-949, 2020.
https://doi.org/10.30534/ijeter/2020/01842020
3. G. Suresh and R. Balasubramanian. **An Ensemble Feature Selection Model using Fast Convergence Ant Colony Optimization Algorithm**, *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 4, pp. 1417-1423, 2020.
https://doi.org/10.30534/ijeter/2020/77842020
4. N. Czink, P. Cera, J. Salo, E. Bonek, J.-P. Nuutinen, and J. Ylitalo. **A framework for automatic clustering of parametric MIMO channel data including path powers**, in *IEEE 64th Veh. Technol. Conf.*, 2006, pp.1-5.
https://doi.org/10.1109/VTCF.2006.35
5. C. Gentile. **Using the kurtosis measure to identify clusters in wireless channel impulse responses**, *IEEE Trans. Antennas Propag.*, vol. 61, no. 6, pp. 3392-3395, 2013.
https://doi.org/10.1109/TAP.2013.2253299
6. B. Li, C. Zhao, H. Zhang, Z. Zhou, and A. Nallanathan. **Efficient and robust cluster identification for ultra-wideband propagations inspired by biological ant colony clustering,** *IEEE Trans. Commun.*, vol. 63, no. 1, pp. 286-300, 2015.
https://doi.org/10.1109/TCOMM.2014.2377120
7. S. Cheng, M.-T. Martinez-Ingles, D.P. Gaillot, J.-M. Molina-Garcia-Pardo, M. Liénard, and P. Degauque. **Performance of a novel automatic identification algorithm for the clustering of radio channel parameters**, *IEEE Access*, vol. 3, pp. 2252-2259, 2015.
https://doi.org/10.1109/ACCESS.2015.2497970
8. R. He, W. Chen, B. Ai, A.F. Molisch, W. Wang, Z. Zhong, J. Yu, and S. Sangodoyin. **On the clustering of radio channel impulse responses using sparsity-based methods**, *IEEE Trans. Antennas Propag.*, vol. 64, no. 6, pp. 2465-2474, 2016.
https://doi.org/10.1109/TAP.2016.2546953
9. R. He, Q. Li, B. Ai, Y. Geng, A. Molisch, V. Kristem, Z. Zhong, and J. Yu. **A kernel-power-density-based algorithm for channel multipath components clustering**, *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7138-7151, 2017.
https://doi.org/10.1109/TWC.2017.2740206
10. Y. Li, J. Zhang, Z. Ma, and Y. Zhang. **Clustering analysis in the wireless propagation channel with a variational Gaussian mixture model**, *IEEE Trans. Big Data*, vol. 6, no. 2, pp. 223-232, 2020.
https://doi.org/10.1109/TBDATA.2018.2840696
11. H. Xue, L. Tian, Y. Zhang, C. Wang, J. Zhang, W. Li, and H. Wei. **A novel clustering method based on density peaks and its validation by channel measurement**, in *Proc. 12th Eur. Conf. Antennas Propag. (EuCAP)*, Apr. 2018, pp. 1-5.
https://doi.org/10.1049/cp.2018.0700
12. D. Abinoja and L. Materum. **Evaluation of Techniques for Identifying Multipath Propagation Clusters in Wireless Systems by Bayesian Information Criterion Model-Based Clustering**, *Lecture Notes Adv. Res. Eng. Inf. Technol. 1*, pp. 97-102, 2019.
13. A. Teologo and L. Materum. **Cluster-Wise Jaccard Accuracy of KPower Means on Multipath Datasets**, *Int. J. Emerg. Trends Eng. Res.*, vol. 7, no. 8, pp. 203-208, 2019.
https://doi.org/10.30534/ijeter/2019/16782019
14. O. Bialer, N. Garnett, and D. Levi. **A Deep Neural Network Approach for Time-Of-Arrival Estimation in Multipath Channels**, in *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2018, pp. 2936-2940.
https://doi.org/10.1109/ICASSP.2018.8461301
15. A. Abdallah and Z. Kassas. **Deep Learning-Aided Spatial Discrimination for Multipath Mitigation**, in *IEEE/ION Pos. Loc. Nav. Symp. (PLANS)*, 2020, pp. 1324-1335.
https://doi.org/10.30534/ijeter/2019/16782019
16. L. Liu, C. Oestges, J. Poutanen, K. Haneda, P.

Vainikainen, F. Quitin, F. Tufvesson, and P. De Doncker. **The COST 2100 MIMO channel**, *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 92-99, 2012. https://doi.org/10.1109/MWC.2012.6393523

17. J. Blanza, A. Teologo, and L. Materum. **Datasets for multipath clustering at 285 MHz and 5.3 GHz bands based on COST 2100 MIMO channel model**, [Online]. Available: http://dx.doi.org/10.21227/4cb9-hf81, 2019.
18. J. Blanza, A. Teologo, and L. Materum. **Datasets for multipath clustering at 285 MHz and 5.3 GHz bands based on COST 2100 MIMO channel model**, in *9th Int. Symp. Multimedia Commun. (ISMAC 2019)*, Aug. 2019, pp. 1-5. https://doi.org/10.1109/ISMAC.2019.8836143
19. M. Kampffmeyer, S. Løkse, F. M. Bianchi, L. Livi, A.-B. Salberg, and R. Jenssen. **Deep divergence-based approach to clustering**, *Neural Netw.*, vol. 113, pp. 91-101, 2019. https://doi.org/10.1016/j.neunet.2019.01.015