IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 11 January 2024, accepted 7 February 2024, date of publication 12 February 2024, date of current version 27 February 2024.
*Digital Object Identifier 10.1109/ACCESS.2024.3365188*

RESEARCH ARTICLE

# WLAN Channel Status Duration Prediction for Audio and Video Services Using Probabilistic Neural Networks

**YAFEI HOU, (Senior Member, IEEE), AND SATOSHI DENNO, (Member, IEEE)**
Faculty of Environmental, Life, Natural Science and Technology, Okayama University, Okayama 700-8530, Japan
Corresponding author: Yafei Hou (yfhou@okayama-u.ac.jp)

This work was supported in part by the GMO Internet Foundation, and in part by the JSPS KAKENHI under Grant 20K04484.

**ABSTRACT** Due to massive increase in wireless access from smartphones, IoT devices, WLAN is aiming to improve its spectrum efficiency (SE) using many technologies. Some interesting techniques for WLAN systems are flexible allocation of frequency resource and cognitive radio (CR) techniques which expect to find more useful spectrum resource by modeling and then predicting of channel status using the captured statistics information of the used spectrum. This paper investigates the prediction accuracy of busy/idle duration of two major wireless services: audio service and video service using neural network based predictor. We first study the statistics distribution of their time-series busy/idle (B/I) duration, and then analyze the predictability of the busy/idle duration based on the predictability theory. Then, we propose a data categorization (DC) method which categorizes the duration of recent B/I duration according the their ranges to make the duration of next data be distributed into several streams. From the predictability analysis of each stream and the prediction performance using the probabilistic neural network (PNN), it can be confirmed that the proposed DC can improve the prediction accuracy of time-series data in partial streams.

**INDEX TERMS** Channel status duration prediction, WLAN audio/video traffic, data predictability analysis, probabilistic neural network (PNN).

## I. INTRODUCTION

The next-generation wireless technology is aiming to hold diversified services, applications considering the usage of huge number of devices which is reshaping and facilitating our daily life [1]. Among them, wireless local area network (WLAN) are widely used in indoor scenarios such as home, office for wireless access which holds over 80% wireless traffic [2]. To meet this trend of massive increase in access and capacity, WLAN with Wi-Fi hot-spots is expected to transmit or receive increasing amounts of data in more indoor scenarios. Therefore, some efficient techniques are prompting to further improve the spectrum efficiency (SE) of WLAN [3]. There are many ongoing research items for realizing the purpose such as increasing the number of antennas to 8 streams for IEEE 802.11ac (Wi-Fi 5) and enlarging the bandwidth to 160MHz band for IEEE 802.11ax (Wi-Fi 6) [4]. In addition, mmWave MIMO technique is considered for IEEE 802.11ay which can utilize the bandwidth of 8 GHz over 60 GHz frequency band [5].

One of the most promising techniques for WLAN system is cognitive radio (CR) technology. CR dynamically estimates the channel status, predicting the coming channel state and finding the promising wireless resources in its vicinity to avoid congestion and reduce interference with a smart way [6]. Recent progresses of CR technique using machine learning (ML) and predictability analysis on spectrum usage have further cultivated and prompted the next generation WLAN system [7]. To achieve high SE, WLAN system with CR needs to correctly detect the channel usage status or data pattern. A realistic, accurate channel occupancy pattern for the wireless traffic of each wireless service is essential and extremely useful. Then, WLAN system also aims to correctly capture current channel state and predict the coming channel

The associate editor coordinating the review of this manuscript and approving it for publication was Wanqing Zhao.

status before its start of packet transmission. A tutorial paper [8] has summarized the existing prediction schemes for the CR optimization and allocation of the wireless frequency resource. These research activities have presented many efficient and smart methods for CR system which mostly consider the prediction for some key statistics parameters such as spectrum usage, channel occupancy or utilization rate with the time resolution like second, hour or day. These predicted statistics parameters and properties would help the improvement of spectrum utilization and the policies for access optimization.

If WLAN system with CR technology can correctly predict its start and end of the channel busy/idle status or status duration in very short time, it can efficiently utilize the available frequency bands in a resource-efficient way, and further improve the SE especially when the limited wireless channel resource is requested by a large number of devices. As one example, CR technology can utilize channel idle duration scattered in multiple WLAN frequency bands by splitting one long transmission packet into several short sub-packets, then transmit them on the multiple bands [9], [10]. In addition, correct prediction within short period will improve the overall SE of all WLAN bands especially for some environments with heavy wireless traffic where the wireless spectrum resources are limited. The prediction of channel status duration is extremely important but difficult and sometime impossible because the time-series status duration is unstable and disordered in short period. This makes the prediction of the status duration to be intangible and difficult [6], [11].

Machine learning (ML)-based prediction methods show powerful capabilities to cope with some complicated real-time problems which are difficult to be described by model-driven approach. By learning the relationship among data, data-driven ML-based method can capture and analyze ambient WLAN signals to understand their physical characteristics for usage. Recently, many ML-based methods have been employed for wireless system, such as massive MIMO channel prediction using deep convolutional neural network (CNN) [12] and multi-layer perceptron (MLP) [13]. In [14], a joint design of beamforming vector and learning rate in MIMO over-the-air computation for federated learning has been researched. In addition, with reduced training overhead and small dataset, a fast adaptive channel prediction technique based on a meta-learning algorithm for massive MIMO communications has been investigated in [15]. Reader can find the details of ML-based methods in [16] which systematically reviewed the current representative "learning to optimize" techniques in diverse domains of 6G wireless networks and investigating the specifically designed ML frameworks from the perspective of optimization. These researches provided many efficient ways to predict the channel status duration of WLAN system.

In addition, after obtaining the statistics of spectrum usage such as busy or idle duration, some fundamental but important questions for the time-series data are some like that: to what level is the data predictable? Whether there exists some methods to increase the predictability of the time-series data? Regarding to these questions, using Fano inequality and statistical entropy measures [17], [18], it can provide some basic analysis on the predictability of the time-series data. In addition, we also proposed a data categorization (DC) method to divide time-series data into several streams and each stream has different distribution property [19] with different level of the predictability. The prediction accuracy of some streams with a large predictability can be improved.

There is little research about the prediction of busy/idle duration for WLAN system. The reason perhaps is that such research is hard to be applied for real system because the operation period for prediction needs to be within the unit of busy/idle duration which is less than millisecond. In addition, due to the short time of busy/idle duration data, the distribution or statistics property of data used for prediction is dramatically changed which generates outliers and deteriorates the prediction accuracy. However, using new machine learning technology and high powered sensing devices, the WLAN system should be considered for hold more device access with smart idle resource management and prediction within more short time scale. On the other hand, although many services generate WLAN traffic, two widely and major used WLAN traffics, audio service and video service, are more than 60% traffic over total IEEE 802.11 WLAN traffic [20]. Therefore, investigating the statistics properties of busy/idle duration of audio and video services and their predictability can benefit the CR technique used for WLAN system [21], [22], [23], [24].

In this paper, using the Fano inequality and statistical entropy measures, we study the modeling and predictability analysis on time-series busy/idle duration of 802.11 WLAN traffic for video and audio services [25]. Then we show their prediction performance using PNN based predictor with the proposed DC method. The major novelties and contributions of this paper are as follows.

1) We study the statistics distribution properties of time-series busy/idle duration of 802.11 WLAN traffic for audio and video services. The similar work considering for VoIP skype service and FTP service has been investigated in [24]. Furthermore, we will consider video and audio services on two scenarios: the ideal case (the data is captured in an anechoic chamber), and real case (the data is captured in an indoor office) to compare the different distribution properties between two cases.
2) The predictability of the time-series busy/idle duration of both services is analyzed to find the lower bound and upper bound of predictability. Although it is still unknown how to realize the bound of accurate prediction, the comparison and analysis show the different predictability of time-series busy/idle duration between ideal and normal cases.

3) We propose the busy/idle (B/I) duration prediction method using PNN based predictor with the DC method. In [19], we have analyzed that, by introducing the categorization range of the duration of recent B/I statuses, the duration of the next B/I status can be distributed into several sets (streams) with different predictability. Then even the low-complexity AR-based predictor can achieve the promising prediction accuracy for some streams with high predictability. In addition, it is known that ML-based probabilistic neural network (PNN) predictor can learn the relationship among the data. Therefore, whether DC is required or not for PNN predictor is unclear. This paper contributes to this topic.

This paper is organized as follows. Section II describes the methodology of data collection and the modeling of the collected busy and idle duration of both services. Section III analyzes the lower bound and upper bound of predictability of the collected busy/idle duration data and shows the predictability results using DC method for the time-series busy and idle duration data. The PNN-based predictor with the DC method and its different prediction performance for different streams are discussed in Section IV. The paper concludes the results in final Section V.

## II. DATA COLLECTION AND MODELING OF TIME-SERIES BUSY AND IDLE DURATION

### A. THE WIRELESS TRAFFIC EXPERIMENTS AND DATA COLLECTION

Several wireless traffic experiments of WLAN based audio and video services have been conducted to obtain the data of the time-series busy and idle duration. The data process is shown in Fig. 1. The experiments of WLAN based video services using two resolutions as 1K P and 480 P, and audio service with average transmission rate as 320 kbps using WLAN system operated at 5 GHz band have been carried out in an indoor environment (real case, Fig. 2a). For comparison, we also captured the time-series data of spectrum usage of the same experiments in an anechoic chamber (ideal case, Fig. 2b) to study the impact of interference from other services in the real case.

Frame header data of packets has been recorded using a commercial sniffing software on Channel 36 in W52 at the 5 GHz band. The data-rate, frame arrival time, and frame length were extracted, and the number of bits per symbol, used bandwidth and the standard of system (IEEE 802.11b/g/n) were also recorded from the data-rate information of captured frames according to the IEEE 802.11-2016 standards [26] which employed orthogonal frequency division multiplexing (OFDM) technique. Then, the duration of frames was estimated from the value of the required number of OFDM symbols and the corresponding duration of MAC header with PHY preamble. Then the busy/idle duration sequence was calculated with a granularity of 9 $\mu$s per slot or point according to the current WLAN standards.

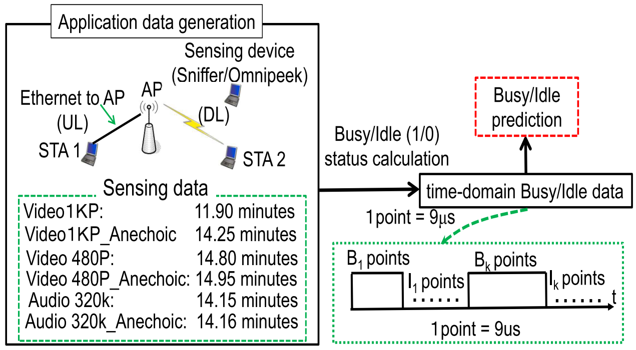
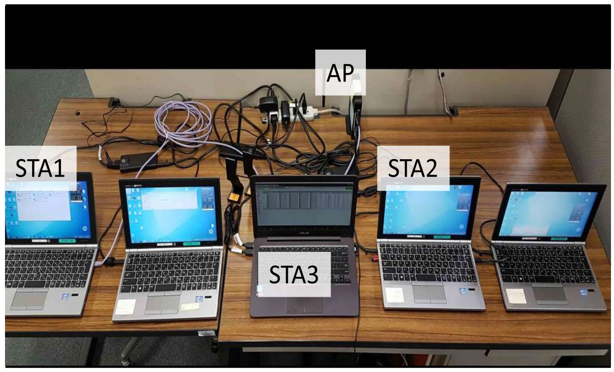
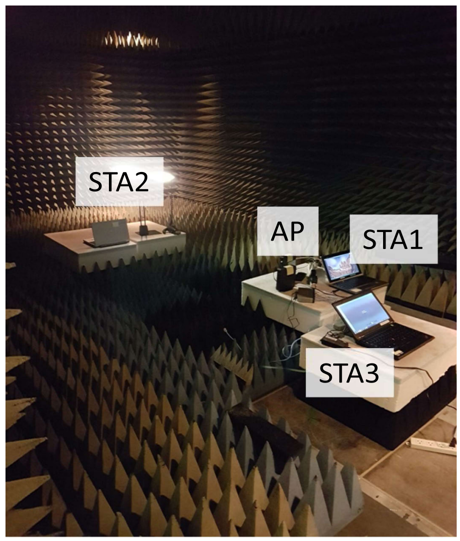



**FIGURE 1.** Main process of experiment.




(a) Office scenario (real case)




(b) Anechoic chamber (ideal case)

**FIGURE 2.** Experiment of data capture.

### B. THE DISTRIBUTION PROPERTY OF BUSY AND IDLE DURATION

For fitting the statistic probability distribution, the empirical cumulative distribution function (CDF) and probability density function (PDF) of collected time-series data need to be calculated. Table 1 lists some major probability distribution models employed in the fitting process, which are exponential (EX), gamma (GM), log-logistic (LLG),

**TABLE 1.** Statistics probability distribution functions for fitting.

| Distribution function | PDF $f(x;\text{parameters})$ | CDF $F(x;\text{parameters})$ |
|---|---|---|
| Exponential (EX) | $f_{EX}(x;\lambda)=\begin{cases}\lambda e^{-\lambda x} & x\geq 0,\\ 0 & x<0.\end{cases}$ | $F_{EX}(x;\lambda)=\begin{cases}1-e^{-\lambda x} & x\geq 0,\\ 0 & x<0.\end{cases}$ |
| Generalized Pareto | $f_{GP}(x;\mu,\sigma,\xi)=\begin{cases}(\frac{\xi(x-\mu)}{\sigma}+1)^{-\frac{\xi+1}{\xi}} & \xi\neq 0,\\ e^{-\frac{(x-\mu)}{\sigma}} & \xi=0.\end{cases}$ | $F_{GP}(x;\mu,\sigma,\xi)=\begin{cases}1-\left(\frac{\sigma+\xi(x-\mu)}{\sigma}\right)^{\frac{-1}{\xi}} & \xi\neq 0,\\ 1-e^{-\frac{(x-\mu)}{\sigma}} & \xi=0.\end{cases}$ |
| Log-normal (LN) | $f_{LN}(x;\mu,\sigma)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right)$ | $F_{LN}(x;\mu,\sigma)=\frac{1}{2}\operatorname{erfc}\left(-\frac{\ln x-\mu}{\sigma\sqrt{2}}\right)$ |
| LogLogistic (LLG) | $f_{LLG}(x;\mu,\sigma)=\frac{(\frac{\sigma}{\mu})(\frac{x}{\mu})^{(\sigma-1)}}{(1+(\frac{x}{\mu})^{\sigma})^2}$ | $F_{LLG}(x;\mu,\sigma)=\frac{(\frac{x}{\mu})^{\sigma}}{1+(\frac{x}{\mu})^{\sigma}}.$ |
| Generalized Extreme Value (GEV) | $s=\frac{x-\mu}{\sigma}$ <br> $f_{GEV}(x;\mu,\sigma,\xi)=\begin{cases}(1+\xi s)^{\frac{-1-\xi}{\xi}}\exp(-(1+\xi s)^{\frac{-1}{\xi}} & \xi\neq 0,\\ \exp(-s)\exp(-\exp(-s)) & \xi=0.\end{cases}$ | $s=\frac{x-\mu}{\sigma}$ <br> $F_{GEV}(x;\mu,\sigma,\xi)=\begin{cases}\exp(-(1+\xi s)^{\frac{-1}{\xi}} & \xi\neq 0,\\ \exp(-\exp(-s)) & \xi=0.\end{cases}$ |
| Gamma (GM) | $f_{GM}(x;\alpha,\beta)=\frac{1}{\beta^{\alpha}\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta}$ | $F_{GM}(x;\alpha,\beta)=\frac{\gamma(\alpha,\beta x)}{\Gamma(\alpha)}$ |
| Inv-Guassian (IG) | $f_{IG}(x;\mu,\lambda)=\sqrt{\frac{\lambda}{2\pi x^3}}exp\left[-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right]$; $\Phi(x)$: Normal distr. | $F_{IG}(x;\mu,\lambda)=\Phi(\sqrt{\frac{\lambda}{x}}(\frac{x}{\mu}-1))+e^{(\frac{2\lambda}{\mu})}\Phi(-\sqrt{\frac{\lambda}{x}}(\frac{x}{\mu}-1))$ |

Log-normal (LN), generalized extreme value (GEV), generalized Pareto (GP) and Inverse Gaussian (Inv-Gaussian) distributions, respectively.

To estimate the major parameters of probability distribution from empirical data, an efficient inference technique which based on maximum likelihood estimation (MLE) is widely adopted. The authors in [27] have utilized the method of moment (MOM) inference scheme to estimate the parameters of statistics distribution and compared with that of using MLE-based method. Their research results concluded that the MLE-based method is generally better than that of the MOM-based scheme. Therefore, MLE-based method is selected to calculate the parameters of different distribution functions. To measure the suitability or accuracy of the fitting results, after fitting, we choose the probability distribution function which has minimum value of Kolmogorov-Smirnov distance between the empirical data and model data generated from the distribution function as the same to that in [27].

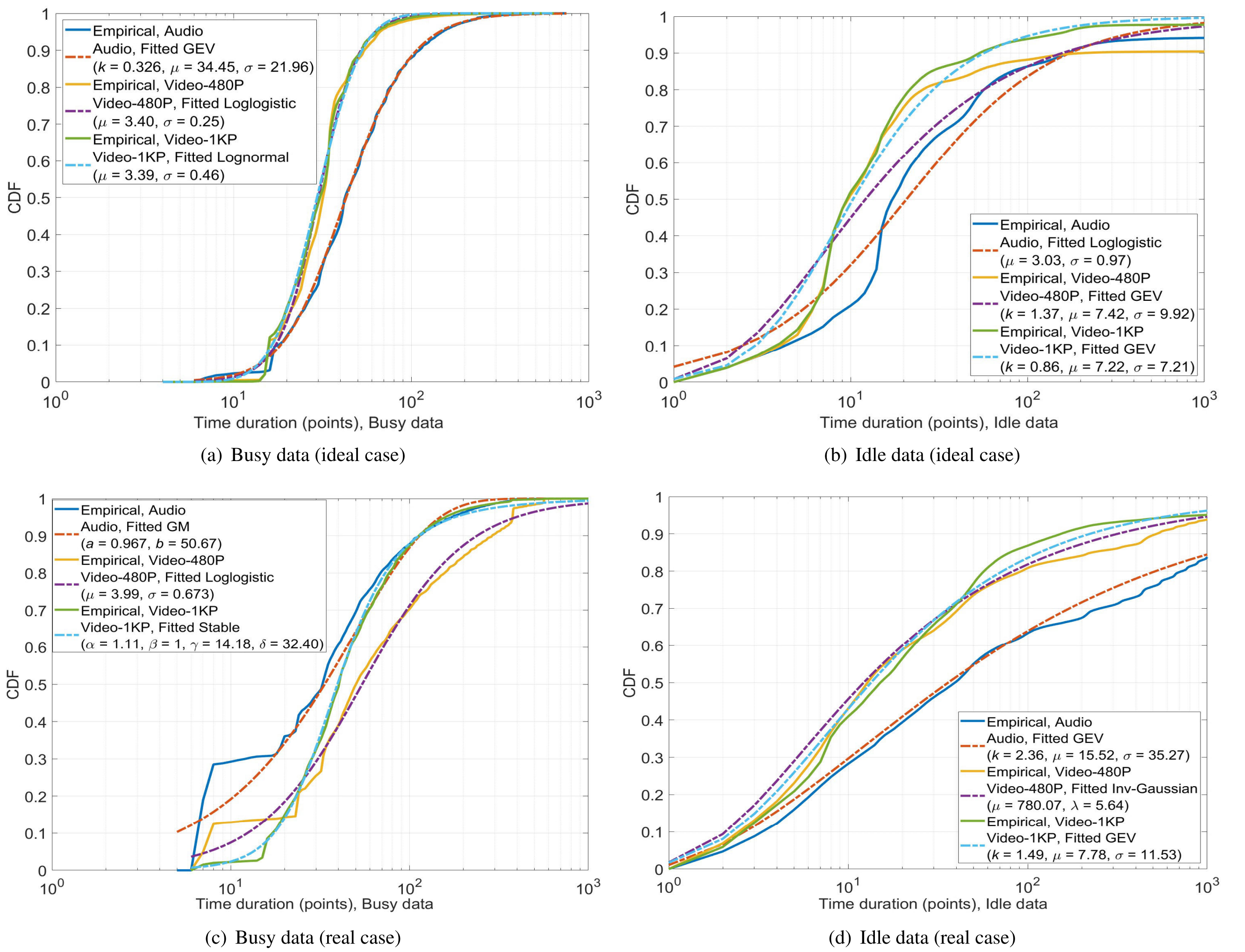
The modeling fitting results of the time-series busy and idle duration collected from the ideal case and real case are given in Fig. 3 respectively. For the modeling fitting results of the ideal case in Fig. 3(a) and (b), it easy to find that the busy duration data is well-fitted with some simple statistics distributions than the idle duration data, especially for the duration data of WLAN-based audio service. In addition, the statistics distribution of the data duration collected from video service has more concentrated on the specific range than that of the data duration collected from audio service. The reason is due to that the data duration of the audio service includes frames with different length. However, the statistics distribution of idle duration of both services appear more randomness which reduces the accuracy of modeling fitting. For real case that some other wireless interference occurs. The results in Fig. 3(c) and (d) showed that the fitting error of statistics distribution increases for the busy duration data of both video and audio service. In addition, idle duration data collected over real case has a better fitting accuracy than that over ideal case. The results imply busy duration of both services over the ideal case has more constant traffic pattern. The idle duration appears more randomness than that of the real case. The randomness makes the fitting of statistics distribution more difficult.

## III. PREDICTABILITY ANALYSIS OF THE TIME-SERIES BUSY IDLE DATA

After obtaining the statistics of the busy or idle duration, it is still difficult to find that the time-series busy/idle data is easy to be predicted or not because their PDF or CDF properties cannot reveal their correlation properties which highly decides the predictability level. Intuitively, for some time-series data like additive white Gaussian noise (AWGN) which has no correlation among the succeeding data, the predictability of these is not larger than $1/M$ ($M$ is the number of different kind of values of time-series data. For AWGN, $M = \infty$). Therefore, some fundamental but important questions for the time-series data are that: to what level is the data predictable? Whether there exists some methods to improve the predictability of the time-series data? Regarding to these questions, using Fano inequality and statistical entropy measures [17], [18], it can provide the basic analysis on the level of the predictability of time-series data. For the second question, we also proposed a data categorization (DC) method to separate the time-series data into several streams according to the setting range, and each stream has different distribution and correlation property [19] which differentiates the level of each stream.

### A. PREDICTABILITY DEFINITION AND ANALYSIS

Here we firstly explain the predictability concept which decides the fundamental prediction limitations or bound of time-series data.



(a) Busy data (ideal case)
(b) Idle data (ideal case)
(c) Busy data (real case)
(d) Idle data (real case)

**FIGURE 3. The statistics distribution fitting for busy/idle duration data.**

One random variate $X$ is assumed with $M$ kind of values. The entropy of $X$ is represented as $S(X)$ as follows

$$S(X) = -\sum_{i=1}^{M} x(i)\log(f(x_i)), \tag{1}$$

with $f(x)$ as the PDF of $X$. In similar way, the average entropy of the time-series data $\mathbf{X}$ ($[x_1, x_2, \ldots, x_N]$) with length as $N$ is represented as

$$S(\mathbf{X}) = \lim_{N\to\infty} \frac{1}{N} S(x_1, x_2, \ldots, x_N). \tag{2}$$

We also use a conditional entropy defined as $S(\mathbf{X}')$ and

$$S(\mathbf{X}') = \lim_{N\to\infty} S(x_N|x_{N-1}, x_{N-2}, \ldots, x_1). \tag{3}$$

The average entropy $S(\mathbf{X})$ is equal to its conditional entropy $S(\mathbf{X}')$ ($N \to \infty$) represented as

$$\begin{aligned} S(\mathbf{X}) &= \lim_{N\to\infty} \frac{1}{N} S(x_1, x_2, \ldots, x_n) \\ &= \lim_{N\to\infty} S(x_N|x_{N-1}, x_{N-2}, \ldots, x_1) \\ &= \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} S(x_i|h_{i-1}) = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} S(i). \end{aligned} \tag{4}$$

Here $S(i)$ and $h_{i-1}$ are $S(i) = S(x_i|h_{i-1})$ and $h_{i-1}$= $[x_{i-1}, x_{i-2}, \ldots, x_1]$, respectively.

The $S(\mathbf{X})$ of $N$-length time-series data is related to the value of $\mathbf{X}$ and their joint PDF, and is difficult to be obtained. Usually, its approximate value is calculated as $S^{Real}(\mathbf{X})$ using Lempel-Ziv algorithm [28].

If there is no correlation among $N$-length time-series data, only using its PDF, the average entropy $S^{Unc}(\mathbf{X})$ can be simply calculated as

$$S^{Unc}(\mathbf{X}) = -\sum_{i=1}^{M} x(i)\log(f(x_i)). \tag{5}$$

For time-series data, we suppose there exists one prediction method and its accuracy probability $P_r$ is given as

$$P_r = \text{Prob}\{\hat{x}_n = x_n|h_{n-1}\}. \tag{6}$$

The predictability value of one time-series data can be defined as the maximum accuracy probability among all prediction

**TABLE 2.** The entropy and predictability probability (ideal case).

| Data type | $S^{Real}$ | $S^{Unc}$ | $\Pi^{Real}$ | $\Pi^{Unc}$ | $M$ |
|---|---|---|---|---|---|
| Busy (Video 480P) | 4.36 | 5.46 | 0.57 | 0.44 | 295 |
| Idle (Video 480P) | 4.94 | 6.19 | 0.67 | 0.57 | 4766 |
| Busy (Video 1KP) | 4.44 | 5.61 | 0.58 | 0.44 | 393 |
| Idle (Video 1KP) | 4.34 | 5.34 | 0.69 | 0.61 | 2301 |
| Busy (Audio 320K) | 5.20 | 6.72 | 0.51 | 0.33 | 404 |
| Idle (Audio 320K) | 5.37 | 6.77 | 0.61 | 0.49 | 2361 |

**TABLE 3.** The entropy and predictability probability (real case).

| Data type | $S^{Real}$ | $S^{Unc}$ | $\Pi^{Real}$ | $\Pi^{Unc}$ | $M$ |
|---|---|---|---|---|---|
| Busy (Video 480P) | 3.88 | 6.97 | 0.68 | 0.35 | 635 |
| Idle (Video 480P) | 5.65 | 7.46 | 0.60 | 0.45 | 3582 |
| Busy (Video 1KP) | 4.69 | 6.58 | 0.59 | 0.38 | 561 |
| Idle (Video 1KP) | 4.79 | 6.07 | 0.66 | 0.56 | 4284 |
| Busy (Audio 320K) | 3.35 | 6.10 | 0.72 | 0.42 | 469 |
| Idle (Audio 320K) | 6.52 | 8.81 | 0.56 | 0.38 | 6670 |

methods represented as

$$\Pi(h_{n-1}) = \sup\{\hat{x}_n = x_n | h_{n-1}\}. \tag{7}$$

Therefore, for one $N$-length time-series data, the average predictability value $\Pi$ is given as

$$\Pi = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \Pi(i), \tag{8}$$

with $\Pi(i) \triangleq f(h_{N-1})\Pi(h_{N-1})$.

Using Fano inequality, the relationship between the entropy of time-series data and its predictability can be built as

$$\begin{aligned} S(x_N|h_{N-1}) &\leq -[p\log_2 p + (1-p)\log_2(1-p)] \\ &\quad + (1-p)\log_2(M-1) \\ &\triangleq S_F(p) = S_F(\Pi(h_{N-1})), \end{aligned} \tag{9}$$

where $p$ is $\Pi(h_{N-1})$. From the equation, the upper bound (UB) and low bound (LB) of predictability are given as

$$\begin{aligned} &S^{Real} \leq S_F(\Pi^{Real}) \\ &= -[\Pi^{Real}\log_2 \Pi^{Real} + (1-\Pi^{Real})\log_2(1-\Pi^{Real})] \\ &+ (1-\Pi^{Real})\log_2(M-1), \end{aligned} \tag{10}$$

$$\begin{aligned} &S^{Unc} \leq S_F(\Pi^{Unc}) \\ &= -[\Pi^{Unc}\log_2 \Pi^{Unc} + (1-\Pi^{Unc})\log_2(1-\Pi^{Unc})] \\ &+ (1-\Pi^{Unc})\log_2(M-1). \end{aligned} \tag{11}$$

Using Eqs.(10) and (11), it is easy to calculate the LB $\Pi^{Unc}$ and UB $\Pi^{Real}$ of predictability with its entropy $S^{Unc}$ and $S^{Real}$. However, $\Pi^{Unc}$ and $\Pi^{Real}$ just measure the level or difficulty of the prediction on time-series data, but they do not mean the actual prediction method exists to realize $\Pi^{Real}$. For some complicated time-series data, it is maybe impossible to find such prediction method.

### B. ENTROPY AND PREDICTABILITY OF BUSY/IDLE DURATION DATA

An iterative Lempel-Ziv algorithm [28] can be employed to calculate the value of $S^{Real}$, and explore the relation between time-series busy/idle duration data. In addition, it can also utilizes the partial data for the calculation of $S^{Real}$ if the process system has limited memory used for huge size of data. On the other hand, $S^{Unc}$ is calculated using the probability distribution information of the busy and idle duration with Eq. (5).

Table 2 and Table 3 provide the estimated values of $S^{Real}$, $S^{Unc}$, $\Pi^{Real}$, $\Pi^{Unc}$ and $M$ of busy and idle duration for the ideal case and real case using the Lempel-Ziv algorithm, respectively. The results reveal that the predictability of time-series duration data collected from the ideal case has the larger value than that from the real case except of the busy duration data of audio service. In addition, the predictability value of the prediction methods using PDF information can only reach to about 30%–50%. While exploring the correlation properties among the data, prediction accuracy can be increased to 50%–70%. In addition, from the values of predictability and entropy, the results accord with the intuition that the smaller the entropy is, the higher predictability the data has. Because the higher entropy usually means the more uncertainty of the data.

### C. PREDICTABILITY ADJUSTMENT WITH DATA CATEGORIZATION (DC)

From Eqs.(10) and (11), it is known that for time-series data, the correlation property highly decides its predictability value. For some time-series data without correlation among the data like additive white Gaussian noise, the $\Pi^{Unc}$ and $\Pi^{Real}$ are the same as $1/M$ to 0. In other words, the difference value of the predictability as $(\Pi^{Real} - \Pi^{Unc})$ can be regarded as the gain from the correlation information of time-series data. However, exploiting the correlation property of data is challenging, difficult and impossible for some cases which related to the huge volume, multiple dimensions and length of the time-series data.

But it exists one simple way called as data categorization to increase the predictability by changing the PDF or CDF of the time-series data. In [19], we showed that the busy /idle duration data captured from the channels over the environments with heavy WLAN traffic has strong correlation. By changing their distribution property, the accuracy of prediction for the partial data can be improved using a low-complexity auto-regressive based predictor.

The time-series busy/idle duration of wireless traffic data usually has correlation property. For some WLAN based services such as video, audio or FTP etc., the data packets or frames have some similar patterns or duration distributions. Therefore, the predictability of partial data can be increased using the correlation property or busy/idle status transition probability. For example, from Eq. (5), if the PDF of the duration data is centralized type with a fixed range, the

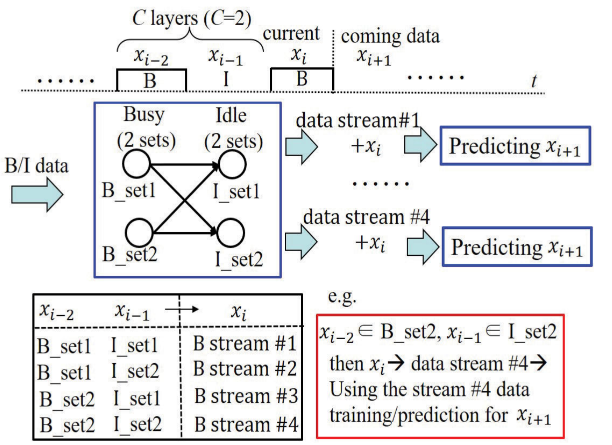



**FIGURE 4.** Data categorization for time-series busy/idle data.

entropy value $S^{Unc}$ becomes small which enlarges the value of predictability $\Pi^{Unc}$.

The idea of DC method for the busy/idle duration data is shown in Fig. 4 as an example. The duration of busy or idle is divided into several non-overlapping sets. For example, The duration data is divided two sets with range as B_set1 $\in [0, a]$ and B_set2 $\in (a, \infty]$ for busy duration $X_{i-2}$ and I_set1 $\in [0, b]$ and I_set2 $\in (b, \infty]$ for idle duration $X_{i-1}$, respectively. Then the busy duration data $X_i$ is categorized according to the set ranges of the busy duration data $X_{i-2}$ and the idle duration data $X_{i-1}$. Since both busy duration data $X_{i-2}$ and idle duration data $X_{i-1}$ are divided into two sets, therefore, according to the set information of previous data $X_{i-2}$ and $X_{i-1}$, the next duration data $X_i$ will be allocated into four different streams with busy data. As an example, if $X_{i-2} \in B\_set2$ and $X_{i-1} \in I\_set2$, then data $X_i$ will be allocated into the fourth duration stream. The similar idea can be further employed for $C$ layers. The captured original busy/idle time-series duration data is firstly allocated according to the categorized set ranges with $C$ layers and each layer includes $S_i$ $(i = 1, \ldots, K)$ sets. Therefore, the busy/idle duration can be divided into $S$ $(S = \prod_{i=1}^{i=C} S_i)$ different streams. After DC process, for each stream, the prediction algorithm is operated for the next duration prediction.

The parameters of the DC method such as $L$, $S_i$ and set ranges decide the distribution properties of the data in each stream. If there is correlation among the time-series data, the DC method will differentiate the predictability of each stream and makes the data in some streams to be easily predicted. However, for some data without any correlation with each other such as AWGN, DC method just makes the data into different stream just like the data sampling and cannot improve the data predictability of any streams. In other words, $\Pi^{Real}$ and $\Pi^{Unc}$ have the same values and no any prediction improvement method exists for this case.

To show the effectiveness of the proposed DC method, here we compare the $S^{Real}$, $S^{Unc}$, $\Pi^{Real}$, $\Pi^{Unc}$ and $M$ value of all busy and idle duration in each stream using DC method.

**TABLE 4.** The predictability results using DC method (Video 480P, real case).

| Data type | $S^{Real}$ | $S^{Unc}$ | $\Pi^{Real}$ | $\Pi^{Unc}$ | $M$ |
|---|---|---|---|---|---|
| Busy (All data) | 3.881 | 6.966 | 0.68 | 0.35 | 635 |
| Busy (Set (1, 1) ) | 5.367 | 6.851 | 0.52 | 0.35 | 540 |
| Busy (Set (1, 2)) | 1.851 | 4.001 | 0.86 | 0.66 | 507 |
| Busy (Set (2, 1)) | 6.383 | 7.862 | 0.41 | 0.23 | 584 |
| Busy (Set (2, 2)) | 6.272 | 7.918 | 0.43 | 0.23 | 605 |
| Idle (All data) | 5.651 | 7.462 | 0.60 | 0.45 | 3582 |
| Idle (Set (1, 1)) | 4.969 | 6.198 | 0.64 | 0.54 | 2700 |
| Idle (Set (1, 2)) | 5.453 | 6.612 | 0.60 | 0.50 | 2304 |
| Idle (Set (2, 1)) | 6.798 | 9.443 | 0.50 | 0.26 | 3076 |
| Idle (Set (2, 2)) | 5.302 | 6.497 | 0.58 | 0.46 | 1173 |

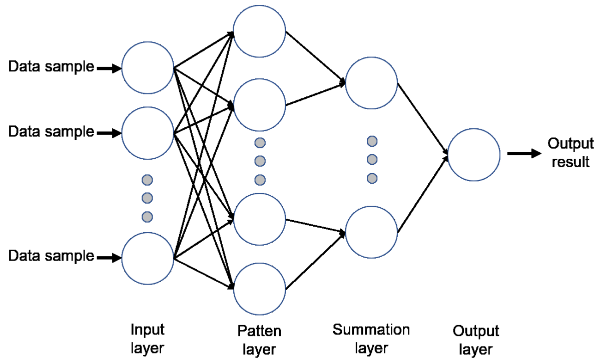



**FIGURE 5.** The architecture of PNN.

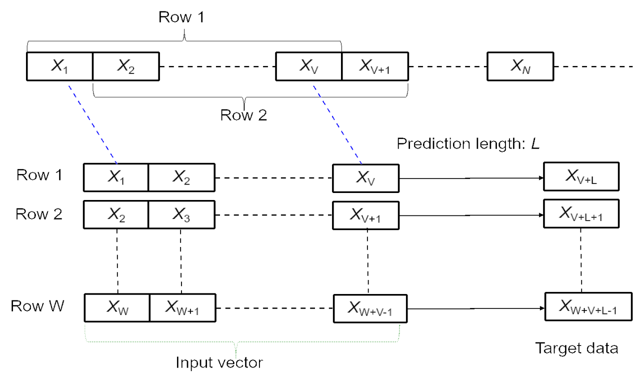
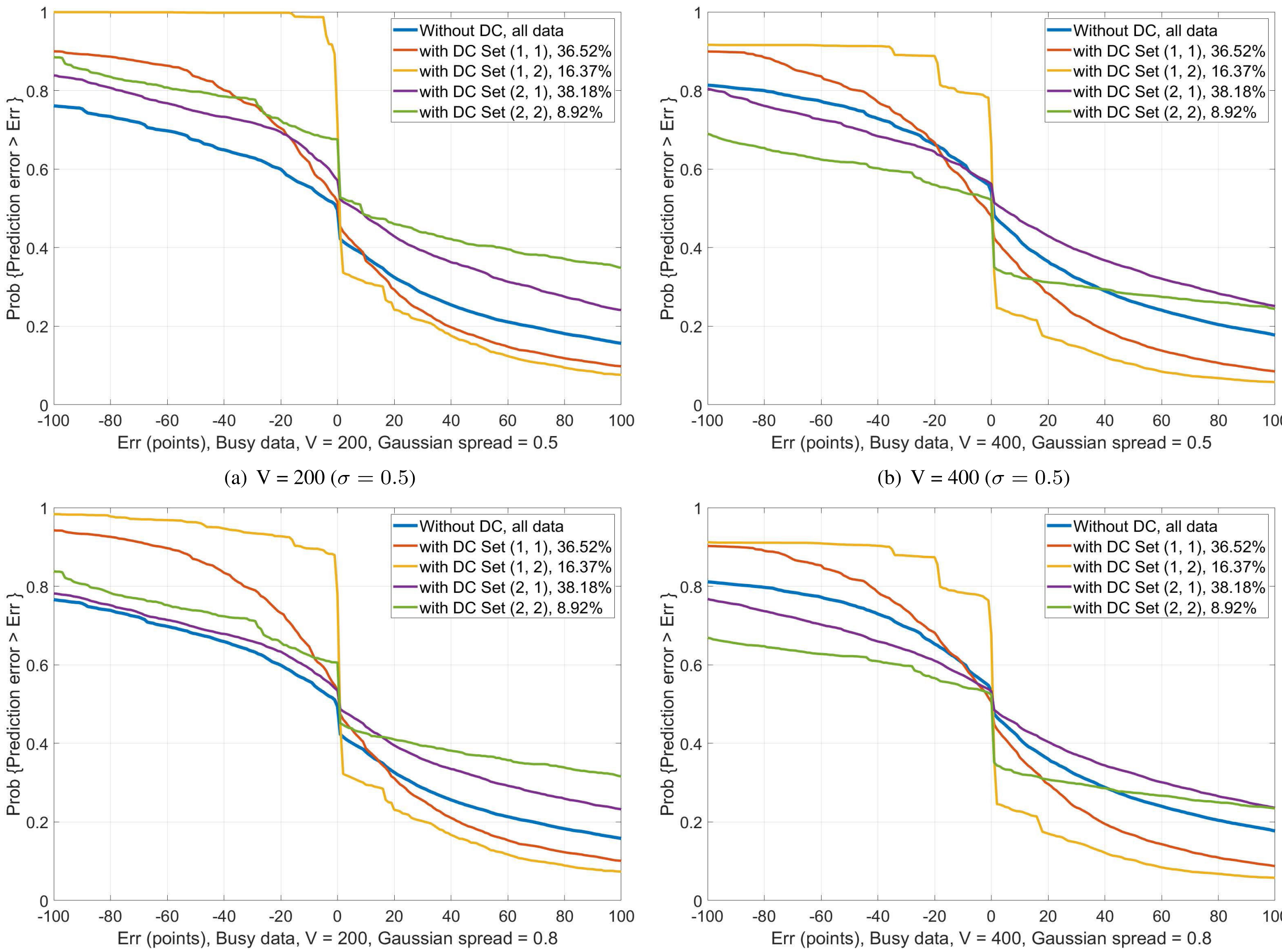



**FIGURE 6.** The structure of input data used in PNN.

We set the number of layer and set as 2 which is the same to Fig. 4. The ranges of two sets are set with $a = b = 0.5$ms for both idle and busy duration.

Table 4 gives the results of $S^{Real}$, $S^{Unc}$, $\Pi^{Real}$, $\Pi^{Unc}$ and $M$ of the busy/idle duration of each stream after the DC process for the video service. It shows that the proposed DC differentiates each stream with different predictability probability. For the busy duration data, the second stream has the larger value $\Pi^{Unc}$ than that of without the proposed DC method. This also means the time-series busy data allocated to second stream is easier to be predicted than other stream. The idle duration data after DC process has similar results. Table 5 shows the calculated results of the busy/idle duration allocated to each stream using DC for the audio service. The




(a) V = 200 ($\sigma = 0.5$)

(b) V = 400 ($\sigma = 0.5$)

(c) V = 200 ($\sigma = 0.8$)

(d) V = 400 ($\sigma = 0.8$)

**FIGURE 7. Busy duration prediction using PNN ($\sigma = 0.5$) with DC method (Video 480P, real case).**

**TABLE 5. The predictability results using DC method (Audio 320K, real case).**

| Data type | $S^{Real}$ | $S^{Unc}$ | $\Pi^{Real}$ | $\Pi^{Unc}$ | $M$ |
|---|---|---|---|---|---|
| Busy (All data) | 3.351 | 6.098 | 0.72 | 0.42 | 469 |
| Busy (Set (1, 1)) | 5.096 | 6.441 | 0.52 | 0.35 | 359 |
| Busy (Set (1, 2)) | 1.993 | 3.756 | 0.84 | 0.67 | 360 |
| Busy (Set (2, 1)) | 5.914 | 7.305 | 0.43 | 0.24 | 392 |
| Busy (Set (2, 2)) | 5.606 | 7.120 | 0.47 | 0.28 | 411 |
| Idle (All data) | 6.519 | 8.814 | 0.56 | 0.38 | 6670 |
| Idle (Set (1, 1)) | 5.812 | 7.463 | 0.59 | 0.45 | 3264 |
| Idle (Set (1, 2)) | 5.651 | 6.961 | 0.56 | 0.43 | 1501 |
| Idle (Set (2, 1)) | 7.640 | 10.305 | 0.45 | 0.20 | 4086 |
| Idle (Set (2, 2)) | 5.623 | 6.954 | 0.53 | 0.39 | 893 |

busy data allocated in second stream is easier to be predicted than other streams, and than that of all busy data without DC method.

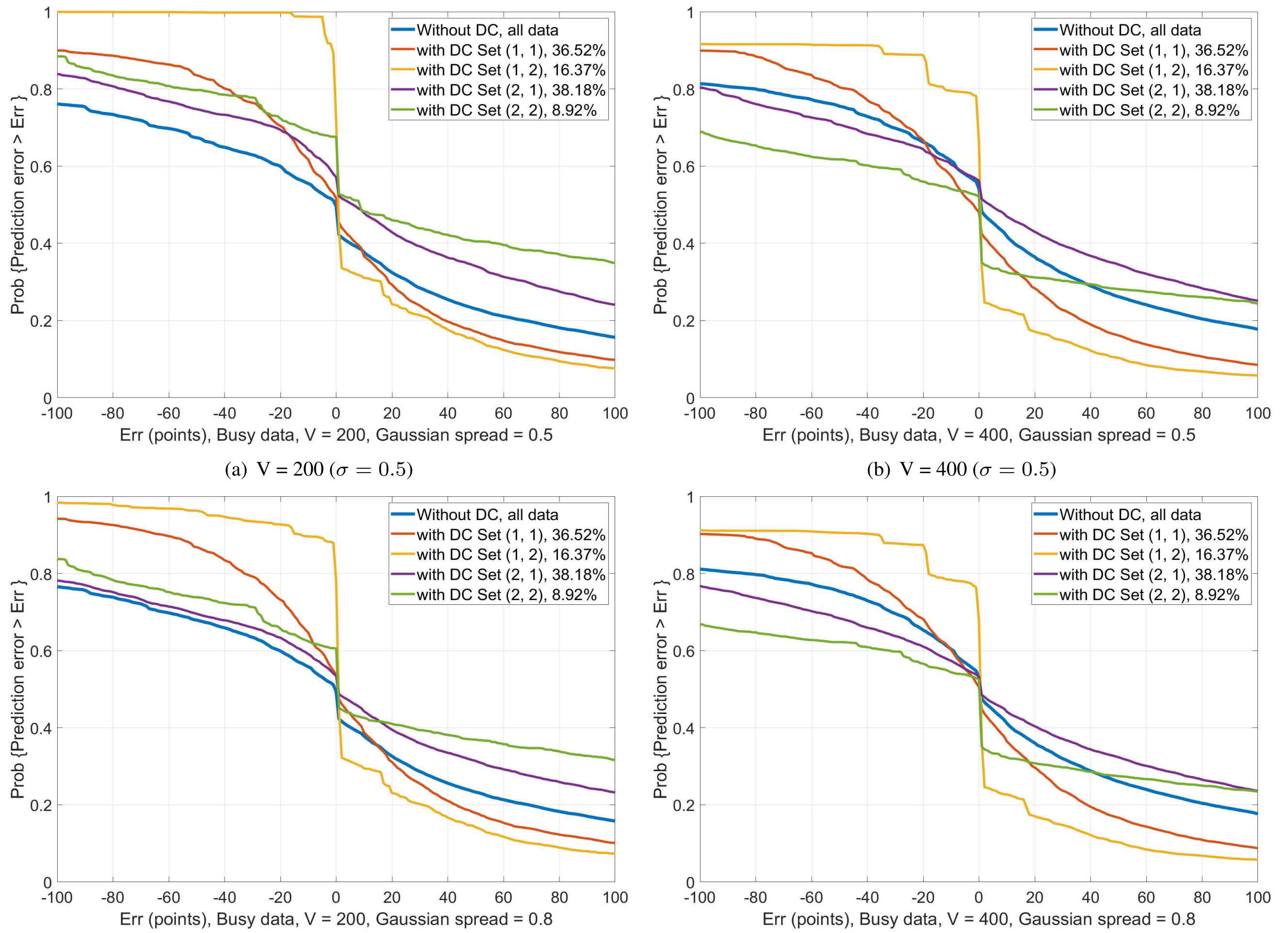
## IV. PROBABILISTIC NEURAL NETWORK (PNN) BASED PREDICTOR WITH DURATION CATEGORIZATION

### A. PROBABILISTIC NEURAL NETWORK

Based on a radial-basis function network, PNN is a developed feed-forward Bayesian network. The core of PNN implements the statistical method called as Kernel Fisher discriminant analysis. It can classify the input training data into different classes by estimating the PDF of each class and optimize the weights of all neurons used for the prediction stage. Fig. 5 shows PNN architecture which includes four layers as input layer, pattern layer, summation layer and output layer, respectively. During training stage, the input vector as training values is received at input layer and then passed to the pattern layer. Here, the dimension of the input vector is same to the number of neurons of input layer. In both pattern and summation layers, the Euclidean distance between the reference vector and the incoming is calculated and the multiplied by the result of a Gaussian activation function at each neuron. Then, the contributions of the $i$-th class is summed to result in a probability from the all neurons in the summation layer which can be represented with following equation as

$$P_i(x) = \frac{1}{\sigma_i\sqrt{2\pi}}\sum_{j=1}^{N_{E_i}} e^{-\frac{\|\mathbf{v}-\mathbf{v}_{i,j}\|^2}{2\sigma_i^2}}. \tag{12}$$

Here vector $\mathbf{v}$ and $\mathbf{v}_{c,j}$ are the sample vector and the $j$th training vector, respectively. We use $N_{E_C}$ and $\sigma_i$ to represent




(a) V = 200 ($\sigma = 0.5$)
(b) V = 400 ($\sigma = 0.5$)
(c) V = 200 ($\sigma = 0.8$)
(d) V = 400 ($\sigma = 0.8$)

**FIGURE 8.** Idle duration prediction using PNN with DC method (Video 480P, real case).

the number of training vectors for $i$-th class and the Gaussian spread, respectively. $||\mathbf{v} - \mathbf{v}_{i,j}||^2$ represents the squared Euclidean distance between the $j$th training vector of $i$-th class and the input vector. Finally the output layer selects the class with the highest probability.

We use Fig. 6 to explain the data structure of the input data used for PNN. Both the training stage and the prediction stage of PNN algorithm have the same structure for the input data. As shown in Fig. 6, the busy or idle data is processed and shifted as data vectors (rows) and then input into the PNN network for the training or the prediction. For $i$-th input row which includes $V$ duration data as $[X_i, \cdots, X_{i+V}$, the target data is represented as $X_{V+L}$ with the prediction length as $L$. For training stage, the value of $X_{V+L}$ is labeled with index as the target of output results of PNN. Therefore, the value of target index is decided by the number of different values or $M$ in Table 2 and Table 3. The PNN method can find the efficient weights of all neurons to build the class relation between the $W \times V$ input matrix and their corresponding $W \times 1$ vector with the target indexes during the training stage. After that, for prediction stage, the index value related to the corresponding duration value is predicted using the weights of all neurons of PNN.

**TABLE 6.** PNN parameters.

| Parameter | Value |
|---|---|
| PNN column length $W$ | 400 |
| PNN rows length $V$ | 200, 400 |
| Step size of sliding-window | 1 sample |
| Number of neurons pattern, summation layer | 400, $M$ |
| Percentage of training samples | 70% |
| Percentage of testing samples | 30% |
| Gaussian spread value $\sigma$ | 0.5, 0.8 |

### B. THE PREDICTION PERFORMANCE OF PROPOSED DC USING PNN

We utilize the busy idle duration data of Video-480P captured at real case for the evaluation of prediction performance of the DC method with PNN. The parameters of the PNN are listed in Table 6. To evaluate the correlation among the time-series data, we set the rows length $V$ as 200 and 400, respectively.

The Gaussian spread $\sigma$ has been set with two values of 0.5 and 0.8 for comparison. We use the CCDF (complementary cumulative distribution function) distribution of prediction error (Prob {prediction error > Err}) of the busy/idle duration prediction in every stream to show the prediction performance of PNN using DC method. Here the prediction error means the difference between the true duration $X_{ture}$ and the predicted value $X_{pred}$ as $(X_{ture} - X_{pre})$. Therefore, the prediction error can be a negative value when $X_{pred} > X_{ture}$.

Figs. 7 show the prediction performance of busy duration using DC method with different rows length $V$, and Gaussian spread $\sigma$ is set as 0.5 and 0.8, respectively. From Fig. 7(a) to (d), the second stream (set (1, 2)) with about 16.373% total busy data shows the higher prediction accuracy than that of using whole data without any DC process. From the values of $S^{Real}$, $S^{Unc}$, $\Pi^{Real}$ and $\Pi^{Unc}$ in Table 4, the values $\Pi^{Real}$ and $\Pi^{Unc}$ of time-series busy data of the second stream are increased to 0.86 and 0.66, respectively. However, the values $\Pi^{Real}$ and $\Pi^{Un}$ of that in other three streams are decreased which means the DC method differentiates the whole data to make partial data to be easily predicted. The PNN predictor can largely improve the prediction accuracy of the second stream compared with that of without DC method. The value $V$ and $\sigma$ also have large impact on the prediction performance of the second stream. The smaller $V$ and $\sigma$ have better prediction accuracy for the second stream. However, for the other three streams, the values of $V$ and $\sigma$ have less influence on the prediction performance for the first stream (set (1,1)) and third stream (set (2, 1)).

Fig. 8 shows the prediction performance of idle duration using DC method with different rows length $V$, and Gaussian spread $\sigma$ is set as 0.5 and 0.8, respectively. Similar to the predicted results of busy duration in Fig. 7, the proposed DC method also differentiates the whole data to make partial data to be easily predicted. However, the prediction accuracy is worse than that of busy duration prediction. The reason is that, different with distribution of idle duration, the busy duration of video service has many fixed length which makes the prediction be more accurate. In addition, the value $\sigma$ of the PNN-based predictor has large impact on the prediction accuracy when the value $V$ is 200.

It should be noted that the performance of prediction accuracy is different with our previous DC method researches in [19] and [25] where we used an auto-regressive based predictor using the duration data captured at one major railway station of Japan [19], and of audio service (real case) [25], respectively. Based on our simulated results, the AR-based predictor has worse prediction accuracy than that of using PNN-based predictor because AR method is generally linear model which cannot represent the complex relationship among time-series data.

## V. CONCLUSION

This paper has investigated the prediction accuracy of busy/idle duration of two wireless services: audio service and video service using neural network based predictor. We first studied the statistics distribution of their time-series busy/idle duration, and then analyzed the predictability of these busy/idle duration based on the predictability theory. Then, we proposed a data categorization (DC) method which categorizes the duration of recent B/I duration according the their ranges to make the duration of next data be distributed into several sets or streams. From the predictability analysis of each stream and the prediction performance using the probabilistic neural network, it can be confirmed that the proposed DC can largely improve the prediction accuracy of time-series data in partial streams.

The proposed DC method still exists many issues need to be researched. For examples, how to decide the best number of sets and set parameters for different time-series data? What is the relationship between categorization data and data predictability? And how to find one low-complexity prediction method to achieve the upper bound of predictability? These issues will be investigated in our future research.

## REFERENCES

[1] A. Kamilaris and A. Pitsillides, "Mobile phone computing and the Internet of Things: A survey," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 885–898, Dec. 2016.

[2] Beyond 5G R&D Promotion Unit, National Institute of Information and Communications Technology, Japan. *Beyond 5G/6G White Paper*. Accessed: Jun. 2023. [Online]. Available: https://beyond5g.nict.go.jp/en/download/index.html

[3] A. Ijaz, L. Zhang, M. Grau, A. Mohamed, S. Vural, A. U. Quddus, M. A. Imran, C. H. Foh, and R. Tafazolli, "Enabling massive IoT in 5G and beyond systems: PHY radio frame design considerations," *IEEE Access*, vol. 4, pp. 3322–3339, 2016.

[4] C. Chen, X. Chen, D. Das, D. Akhmetov, and C. Cordeiro, "Overview and performance evaluation of Wi-Fi 7," *IEEE Commun. Standards Mag.*, vol. 6, no. 2, pp. 12–18, Jun. 2022.

[5] K. Aldubaikhy, W. Wu, N. Zhang, N. Cheng, and X. Shen, "mmWave IEEE 802.11ay for 5G fixed wireless access," *IEEE Wireless Commun.*, vol. 27, no. 2, pp. 88–95, Apr. 2020.

[6] R. Struzak, "Cognitive radio, spectrum, and evolutionary heuristics," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 166–171, Jun. 2018.

[7] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 98–105, Apr. 2017.

[8] O. Simeone, "A very brief introduction to machine learning with applications to communication systems," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 4, pp. 648–664, Dec. 2018.

[9] N. Egashira, K. Yano, S. Tsukamoto, J. Webber, M. Sutoh, Y. Amezawa, and T. Kumagai, "Integrated synchronization scheme for WLAN systems employing multiband simultaneous transmission," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–5.

[10] K. Yano, N. Egashira, J. Webber, M. Usui, and Y. Suzuki, "Achievable throughput of multiband wireless LAN using simultaneous transmission over multiple primary channels assisted by idle length prediction based on PNN," in *Proc. 1st Int. Conf. Artif. Intell. Info. Commun.*, Okinawa, Japan, Feb. 2019, pp. 22–27.

[11] J. Webber, A. Mehbodniya, Y. Hou, K. Yano, and T. Kumagai, "Study on idle slot availability prediction for WLAN using a probabilistic neural network," in *Proc. 23rd Asia–Pacific Conf. Commun. (APCC)*, Dec. 2017, pp. 1–6.

[12] P. Dong, H. Zhang, G. Y. Li, N. NaderiAlizadeh, and I. S. Gaspar, "Deep CNN for wideband mmWave massive MIMO channel estimation using frequency correlation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 4529–4533.

[13] H. Kim, S. Kim, H. Lee, C. Jang, Y. Choi, and J. Choi, "Massive MIMO channel prediction: Kalman filtering vs. machine learning," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 518–528, Jan. 2021.

[14] M. Kim and D. Park, "Joint beamforming and learning rate optimization for over-the-air federated learning," *IEEE Trans. Veh. Technol.*, vol. 72, no. 10, pp. 13706–13711, Oct. 2023.

[15] H. Kim, J. Choi, and D. J. Love, "Massive MIMO channel prediction via meta-learning and deep denoising: Is a small dataset enough?" *IEEE Trans. Wireless Commun.*, vol. 22, no. 12, pp. 9278–9290, Dec. 2023.

[16] Y. Shi, L. Lian, Y. Shi, Z. Wang, Y. Zhou, L. Fu, L. Bai, J. Zhang, and W. Zhang, "Machine learning for large-scale optimization in 6G wireless networks," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 4, pp. 2088–2132, 2023.

[17] G. Ding, J. Wang, Q. Wu, Y.-D. Yao, R. Li, H. Zhang, and Y. Zou, "On the limits of predictability in real-world radio spectrum state dynamics: From entropy theory to 5G spectrum sharing," *IEEE Commun. Mag.*, vol. 53, no. 7, pp. 178–183, Jul. 2015.

[18] J. Sun, L. Shen, G. Ding, R. Li, and Q. Wu, "Predictability analysis of spectrum state evolution: Performance bounds and real-world data analytics," *IEEE Access*, vol. 5, pp. 22760–22774, 2017.

[19] Y. Hou, J. Webber, K. Yano, S. Kawasaki, S. Denno, and Y. Suzuki, "Modeling and predictability analysis on channel spectrum status over heavy wireless LAN traffic environment," *IEEE Access*, vol. 9, pp. 85795–85812, 2021.

[20] H. I. Zawia, R. Hassan, and D. P. Dahnil, "A survey of medium access mechanisms for providing robust audio video streaming in IEEE 802.11aa standard," *IEEE Access*, vol. 6, pp. 27690–27705, 2018.

[21] Cisco. *VNI Complete Forecast Highlights*. [Online]. Available: https://www.cisco.com/c/dam/m/en_us/solutions/service-provider/vni-forecast-highlights/pdf/Global_Business_Highlights.pdf

[22] F. Gringoli, P. Serrano, I. Ucar, N. Facchi, and A. Azcorra, "Experimental QoE evaluation of multicast video delivery over IEEE 802.11aa WLANs," *IEEE Trans. Mobile Comput.*, vol. 18, no. 11, pp. 2549–2561, Nov. 2019.

[23] A. Boujnoui, L. O. Barbosa, and A. Haqiq, "Performance evaluation and tuning of an IEEE 802.11 audio video multicast collision prevention mechanism," *Wireless Netw.*, vol. 26, no. 7, pp. 5047–5061, Oct. 2020.

[24] A. Gupta, S. Agarwal, and S. De, "A new spectrum occupancy model for 802.11 WLAN traffic," *IEEE Commun. Lett.*, vol. 20, no. 12, pp. 2550–2553, Dec. 2016.

[25] Y. Hou, S. Kawasaki, and S. Denno, "Busy/idle duration prediction for video and audio WLAN traffics using autoregressive predictor with data categorization," in *Proc. 23rd Int. Conf. Adv. Commun. Technol. (ICACT)*, Feb. 2021, pp. 259–264.

[26] *802.11-2016 - IEEE Standard for Information Technology—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Standard 802.11-2016, 2016.

[27] L. B. Miguel and F. Casadevall, "Time-dimension models of spectrum usage for the analysis, design, and simulation of cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 2091–2104, Jun. 2013.

[28] I. Kontoyiannis, P. H. Algoet, Y. M. Suhov, and A. J. Wyner, "Nonparametric entropy estimation for stationary processes and random fields, with applications to English text," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1319–1327, May 1998.

[29] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications*. Cham, Switzerland: Springer, 2010.

**YAFEI HOU** (Senior Member, IEEE) received the dual Ph.D. degree from the Kochi University of Technology (KUT), Japan, and Fudan University, China, in 2007. From August 2007 to September 2010, he was a Postdoctoral Research Fellow with Ryukoku University, a Research Scientist with Wave Engineering Laboratories, ATR Institute International, Japan, in March 2014, and a Guest Research Scientist, since October 2016. He was an Assistant Professor with the Nara Institute of Science and Technology. He is currently an Associate Professor with the Faculty of Environmental, Life, Natural Science and Technology, Okayama University, Japan. His research interests include wireless communication systems, and signal processing. He is a Senior Member of IEICE. He received the IEICE Communications Society Best Paper Award, in 2016 and 2020, and the Best Tutorial Paper Award, in 2017. He serves as an Associate Editor for IEEE Access and an Editor for the IEEE Consumer Technology Society News on Consumer Technology (CTSoc-NCT).

**SATOSHI DENNO** (Member, IEEE) received the M.E. and Ph.D. degrees from Kyoto University, Kyoto, Japan, in 1988 and 2000, respectively. He joined NTT Radio Communications Systems Laboratories, Japan, in 1988. He was seconded to ATR Adaptive Communications Research Laboratories, Kyoto, in 1997. From 2000 to 2002, he was with NTT DoCoMo, Japan. He moved to DoCoMo Communications Laboratories Europe GmbH, Germany. He worked as an Associate Professor with Kyoto University, from 2004 to 2011. Since 2011, he has been a Full Professor with the Faculty of Environmental, Life, Natural Science and Technology, Okayama University. He has been engaged in the research and development of digital mobile radio communications. In particular, he has considerable interests in array signal processing, space time codes, spatial multiplexing, and multimode reception. He received the Excellent Paper Award from the IEICE, in 1995, and the Best Paper Award from IEICE Communications Society, in 2020.