Supervised Learning for Performance Prediction in Underwater Acoustic Communications

Evan Lucas and Zhaohui Wang
Dept. of Electrical and Computer Engineering, Michigan Technological University, Houghton, MI 49931, USA

Global Oceans 2020: Singapore - U.S. Gulf Coast | 978-1-7281-5446-6/20/$31.00 ©2020 IEEE | DOI: 10.1109/IEEECONF38699.2020.9389191

***Abstract*—The propagation of acoustic waves under water is a highly complex and stochastic process. Such channel dynamics renders large performance variation in underwater acoustic (UWA) communications. Prediction of the UWA communication performance is critical for selection and adaptation of the communication strategies. This work explores the use of supervised learning for performance prediction in UWA communications. This work first quantifies the transmitter design, the UWA channel characteristics and the receiver design by numerical and categorical parameters. For a chosen performance metric (e.g., the bit error rate or the packet error rate), the performance prediction is cast individually into a numerical prediction problem and a classification problem. Using the data sets from two field experiments, the performance of typical supervised learning methods are examined. The data processing results reveal that some supervised learning methods can achieve fairly good numerical prediction or classification performance, and the discriminative models typically outperform the generative models.**

***Index Terms*—Performance prediction, supervised learning, numerical prediction, classification, underwater acoustic communications**

I. Introduction

The propagation of acoustic waves in the underwater environment is a highly complex stochastic process. For underwater acoustic (UWA) communications, establishing a theoretical relationship among the transmitter design, the channel characteristics and the receiver decoding performance, is a nontrivial task. In this work, we explore the use of different supervised learning methods for the communication performance prediction. The obtained model can guide the selection and adaptation of strategies for point-to-point communications and networking [1], [2].

There are some relevant works in literature. In the setting of wireless sensor networks, supervised learning schemes were introduced in [3] to predict the bit-error-rate (BER) performance, which serves as inputs for the routing decision making. There were several recent attempts for performance prediction in the underwater research area. A logistic regression method was used in [4] to estimate the packet success rate (PSR) based on environmental factors. For adaptive modulation, a boosted decision tree method was introduced in [2] to estimate the BER of different transmission schemes based on the UWA channel parameters. It noted, as did in [5], that using the minimum possible transmission power can reduce the risk of adversaries eavesdropping on transmissions, which is one benefit of reducing transmission power when possible.

This work was supported by the NSF grant ECCS-1651135 (CAREER).

This work conducts a thorough study on supervised learning for the performance prediction in UWA communications. It is cast individually as a numerical prediction problem and as a classification problem. Two field experimental data sets are tested, including one data set collected from the Keweenaw Waterway in August 2014 (abbreviated as KWAUG14) [6], and the other data set from the SPACE08 experiment held off of Martha's Vineyard in 2008 [7]. Both experiments were in shallow water and the transmission waveform was modulated by the orthogonal frequency-division multiplexing (OFDM) technique. With those data sets, the performance of a variety of supervised learning schemes are examined, and recommendations are made.

II. Problem Statement

The UWA communication performance depends on the transmitter design, the channel characteristics and the receiver design. Depending on applications, the performance metric could be the bit error rate, the symbol error rate, or the packet error rate. The goal of performance modeling is to build a model that maps the transmitter design, the channel characteristics and the receiver design to the chosen performance metric.

The key to the performance modeling lies in the parameterization of the transmitter, the channel and the receiver; and how to quantify them by numerical parameters and categorical parameters. In this work, we assume fixed hardware implementation of the transmitter and the receiver, and mainly focus on the software algorithms, including the transmission schemes and the receiver processing methods.

The transmitter is described by categorical parameters including the modulation type (e.g., BPSK or QPSK), the transmission strategy (e.g., single-carrier, multi-carrier or frequency-shift keying) and the error correction coding methods; and the numerical parameters include the transmission power, the pilot overhead, the coding rate, and transmission strategy-specific parameters. The channel is characterized by several commonly used channel parameters including the channel gain-to-noise ratio, the channel delay spread, and the fast fading statistics [6]. Weather and environmental conditions can also be included as channel parameters. The receiver is characterized by the number of receiving channels and the data processing algorithm. The data processing algorithm can be quantified as categorical parameters (e.g., non-iterative or iterative, interference-aware or interference-ignorant). The design parameters in the selected data processing algorithm

Authorized licensed use limited to: Michigan Technological University. Downloaded on December 20,2022 at 17:00:29 UTC from IEEE Xplore. Restrictions apply.
978-1-7281-5446-6/20/$31.00 ©2020 IEEE

can be included as numerical parameters (e.g., the number of iterations in the iterative receiver).

With the above parameterization, the performance modeling naturally becomes a numerical prediction problem. Furthermore, for engineering purposes it may be sufficient to only determine the range of the communication performance. This makes the performance modeling a classification problem.

III. Supervised Learning Methods for Performance Prediction

In this work, several supervised learning methods [8] are examined for numerical prediction and classification. They are briefly described in the following.

A. Regression Methods

Several models are considered and compared for their abilities for performance prediction. Some models are chosen for their simplicity, while others are selected for their capabilities. For example, the "dummy" or "zero rule" method is included to show the bound achieved by always predicting the same value.

- *Linear regression:* A well-known numerical approach; each feature contributes linearly to the predicted output.
- *Linear regression with approximate Radial Basis Function sampling:* Using the Random Kitchen Sinks [9] algorithm to efficiently approximate a radial basis function (RBF) sampling, which transforms the data into a different space.
- *Neural network:* A series of connected neurons contribute to each other through learned weights. In this work, two different neural network frameworks are considered, one with a single layer of ten neurons and the other with two layers of ten neurons.
- *K-Nearest Neighbors:* The concept of K-Nearest Neighbors (KNN) is to use the data as the model. The output value is assigned as the interpolation of the K nearest data points in the training set. The value of K is taken as 9 in this work. Increasing the value of K acts as a form of regularization.
- *Dummy or Zero Rule (ZeroR):* The dummy regression method simply predicts the same value every time. In this work, the median value of the training set is used. This provides an engineering performance lower bound: if a model has performance lower than this, it should probably not be considered.
- *Decision tree:* A graph is created where each node splits into edges based on feature values. There are a variety of algorithms that can be utilized. The Classification and Regression Tree (CART) algorithm [10] implemented in Scikit-Learn [11] is used in this work.

To evaluate the performance of the regression methods, two metrics are considered, including the normalized root mean squared error (NRMSE) and the Pearson correlation coefficient [12] (generally referred to as correlation coefficient) between the true sample values and the predicted sample values. A coefficient of 1 here means that the prediction is linearly related to the actual values. For the dummy method that always guesses a constant, the correlation coefficient is undefined due to the zero variance.

B. Classification Methods

In some applications, it may not be necessary to predict the value of the selected performance metric, but rather estimate the range that the value will lie within. For example one could segment the BER values at the limit for successful complete error correction. This transforms the regression problem into a classification problem.

- *Logistic regression:* A common binary classifier; each data point is assigned a probability of being within a given class. It can be easily extended to a multi-class case by using a one-vs-all approach, where each class is considered separately and the given class is the one with the highest probability.
- *Neural network:* The basic structure used is identical to the one used for numerical regression. Two different neural network frameworks are considered, one with a single layer of ten neurons and the other with two layers of ten neurons. For classification, the output consists of the likelihood of each class. The class with the maximal likelihood is selected for each set of inputs.
- *K-Nearest Neighbors:* Similarly to regression, the training data provides the model. The output value is assigned as the most common value of the K-nearest data points in the training set.
- *Dummy or ZeroR:* The dummy classification method is essentially the same as the dummy regression method. The method used in this work guesses classes using the distribution of the classes in the training data. A simpler approach is to guess the same class every time, for unbalanced classes this can give deceivingly high true positive rates.
- *Decision rule:* A series of rules based on feature values, similar in structure to an IF/ELSEIF statement.
- *Decision tree:* Similar to the regression decision tree, the CART algorithm is used for classification.
- *Naive Bayesian network:* The model attempts to use a chain of learned probabilities to make a classification. The naive Bayesian network is used in this work. Although it makes the assumption that each feature is independent, this classifier is appealing due to it's simplicity and computational efficiency.
- *Random forest:* A random forest is a collection of random trees. The power of the random forest is that each tree is slightly different and decisions about classification are made as an ensemble.

Three different performance metrics are used to compare the classifiers, including the true positive rate (TPR), the false positive rate (FPR) and precision. In the binary classification problem with two classes (class 0 and class 1), class 0 is also referred as the negative class and class 1 is also referred as the positive class. The TPR is defined as the percentage of data points in class 1 that are correctly identified (i.e., the probability of detection). The FPR is defined as the percentage

Authorized licensed use limited to: Michigan Technological University. Downloaded on December 20,2022 at 17:00:29 UTC from IEEE Xplore. Restrictions apply.

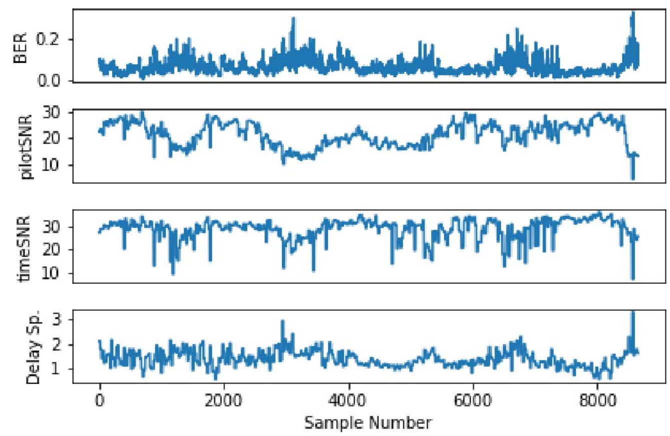
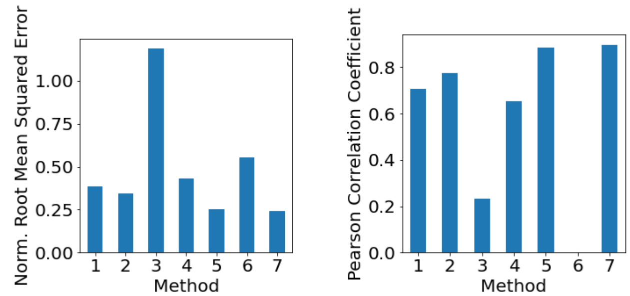
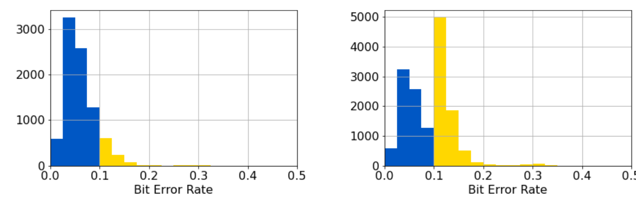


Fig. 1. KWAUG14: Temporal evolution of the BER sequence and the channel statistics. The SNRs are in decible (dB); and the unit of the RMS delay spread is milliseconds.



(a) (b)

Fig. 2. KWAUG14: Numerical prediction performance. Method 1: linear regression; method 2: RBF linear regression; method 3: neural net with 1 layer; method 4: neural net with 2 layers; method 5: KNN; method 6: ZeroR; and method 7: decision tree.



(a) (b)

Fig. 3. KWAUG14: (a) BER distribution and original BER classes with classes color coded. The low BER class consists of 7692 points and the high BER class contains of 968 points. (b) BER classes after SMOTE. Class size for the low BER points remains 7692 and SMOTE increases the number of high BER points to 7692 as well.

of data points in class 0 that are incorrectly identified as class 1 (i.e., the probability of false alarm). Precision is defined as the number of true positives (class 1) divided by the total number of positives classified (correctly or incorrectly). A value of 1 is desirable, as it indicates that everything classified as a positive is correct.

For classification, there exists a well-known problem called the unbalanced class problem, where certain classes do not occur often. An oversampling technique called SMOTE [13], which stands for synthetic minority oversampling technique, can be applied. It uses linear combinations of existing data points to generate new data points and increase the populations of the minority classes.

IV. Experimental Data Processing: KWAUG14

This data set is from a static experiment conducted in the Keweenaw Waterway, August 2014 (abbreviated as KWAUG14) [6]. The experiment lasted about 4.5 days. A waveform of 8.8 seconds in the frequency band $[14, 20]$ kHz was transmitted with a fixed power every 15 minutes over a link of 312 meters. The waveform is modulated by OFDM with a QPSK constellation. Each transmission consists of 20 OFDM blocks. Each OFDM block has 1024 subcarriers with a quarter being pilot subcarriers and 96 being null subcarriers.

The BER is taken as the performance metric. The channel parameters used include:

- *The time-domain SNR:* the received signal-power-to-the-noise-power ratio (SNR) in the time domain;
- *The pilot SNR:* the ratio of the received power on the pilot subcacarriers to that on the null subcarrers, measured in the frequency domain;
- *The channel root-mean-square (RMS) delay spread:* the channel dispersion in delay, weighted by the path amplitudes [6].

In total, there are 8600 data points. Fig. 1 depicts the evolution of the BER sequence and the channel features in time. The data is split into training and test sets, with 80% of data for training and 20% of data for testing.

A. Regression Results

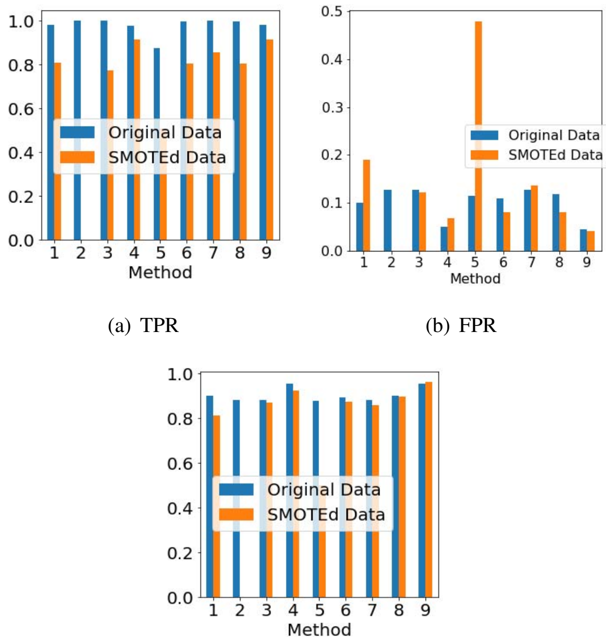
The performance of several regression methods are shown in Fig. 2. The decision tree and the KNN achieve the best performance, with the NRMSE value around 0.25. The performance of linear regression, the RBF transformed linear regression and the neural network with 1 layer are similar. As expected, ZeroR has the worst performance.

B. Classification Results

The data set is segmented into two classes according to the BER value. Because a low BER is preferable, the data points with a BER below 0.1 are assigned as class 1. The other data points are assigned as class 0. Due to the nature of the channel, the vast majority of points are clustered in the lower BER class (i.e., class 1). The SMOTE technique is used to address the unbalanced class problem. The class populations before and after SMOTE can be seen in Fig. 3.

The performance of several classification methods without and with the SMOTE technique are shown in Fig. 4. One can observe that the KNN approach has the best performance (based on precision) for the original data, followed closely by logistic regression, the random forest, and both neural networks. Once SMOTE is applied, an interesting effect is observed. The performance of the single-layer neural network degrades significantly. This is most likely due to overfitting of the model causing the model to only classify points as high BER (i.e., class 0 or a negative). The primary benefit of

Authorized licensed use limited to: Michigan Technological University. Downloaded on December 20,2022 at 17:00:29 UTC from IEEE Xplore. Restrictions apply.



(a) TPR

(b) FPR

(c) Precision

Fig. 4. KWAUG14: Classification performance. Method 1: logistic regression; method 2: neural net with 1 layer; method 3: neural net with 2 layers; method 4: KNN; method 5: ZeroR; method 6: decision rule; method 7: decision tree; method 8: naive Bayesian network; and method 9: random forest.

SMOTE can be seen in the FPR results. The decision becomes less biased towards the low BER class (i.e., class 1 or positive) for the decision rule, decision tree, naive Bayesian network, and the random forest. Precision is boosted for the random forest with SMOTE, although slight degradation is observed for most other classifiers.

V. Experimental Data Processing: SPACE08

A second data set used to evaluate the regression and classification methods is from the Surface Processes and Acoustic Communication Experiment (SPACE08). This experiment was conducted off the coast of Martha's Vineyard at the Air-Sea Interaction Tower, operated by Woods Hole Oceanographic Institution from Oct. 14 to Nov. 1, 2008 [7]. The water depth was about 15 meters. Six bottom adjacent receiving arrays were placed on two paths 90 degrees apart, with three arrays on each path and located at 60, 200, and 1000 meters from the source. The three arrays on one path from near to far are labeled as S1, S3, and S5, respectively. The other three arrays on the other path from near to far are labeled as S2, S4, and S6, respectively. Arrays S1 and S2 both had 48 hydrophone elements, S3 and S4 had 24 hydrophone elements, and S5 and S6 had 12 hydrophone elements. During the experiment, an OFDM-modulated communication waveform within the frequency band $[8, 18]$ kHz was transmitted every two hours. In the transmission waveform, three constellations are included (QPSK, 8-QAM, and 16-QAM), and there are 20 OFDM blocks for each constellation. Each OFDM block has 1024 subcarriers, which are specifically designed with 128 being pilot subcarriers, 384 being data subcarriers and the rest being null subcarriers (please see [14] for the detailed subcarrier distribution).

TABLE I
SPACE08: Total numbers of data points

Receiver	S1	S2	S3	S4	S5	S6
QPSK	42240	37079	28312	28310	19879	20347
8-QAM	42158	37050	28319	28307	19902	20355
16-QAM	42255	36896	28319	28289	19870	20311

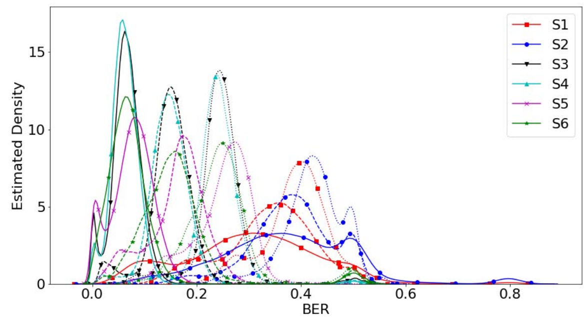


Fig. 5. SPACE08: BER histogram at each receiver and for each constellation. Solid line: QPSK; dashed line: 8-QAM; dotted line: 16-QAM.

In this work, the waveform recorded by each hydrophone and for each constellation is decoded separately, without performing the multi-channel combining. The BER is taken as the performance metric. The same channel parameters as in KWAUG14 are used. Table I lists the total number of data points at each receiver and for each constellation. The BER distribution for each receiver and each constellation is depicted in Fig. 5. A crossplot in Fig. 6 illustrates the relationship among the BER and the channel features. The time-domain SNR is computed based on the estimated channel impulse response.

Similar to KWAUG14, the data is split into training and test sets, with 80% of data for training and 20% of data for testing.

A. Numerical Prediction Results

The performance of several regression methods are shown in Fig. 7. Similarly to the KWAUG14 data set, the decision tree and the KNN give the best results. The neural networks also perform well, as does the RBF transformed linear regression. The ZeroR method shows a range of errors across the data sets that span the performance range of most other methods, giving better results when data sets are more tightly clustered.

B. Classification Results

As shown in Fig. 5, the BER distribution varies across the receivers and the constellation types, and for some data sets, the BER distributions do not overlap. Therefore it is impossible to segment the data using a common BER value for all the data sets. In this work, the data at each receiver and for each constellation is segmented at the median BER value for that given data set. Values below the median BER are considered desirable and assigned as class 1, values above the median are assigned as class 0.

Authorized licensed use limited to: Michigan Technological University. Downloaded on December 20,2022 at 17:00:29 UTC from IEEE Xplore. Restrictions apply.

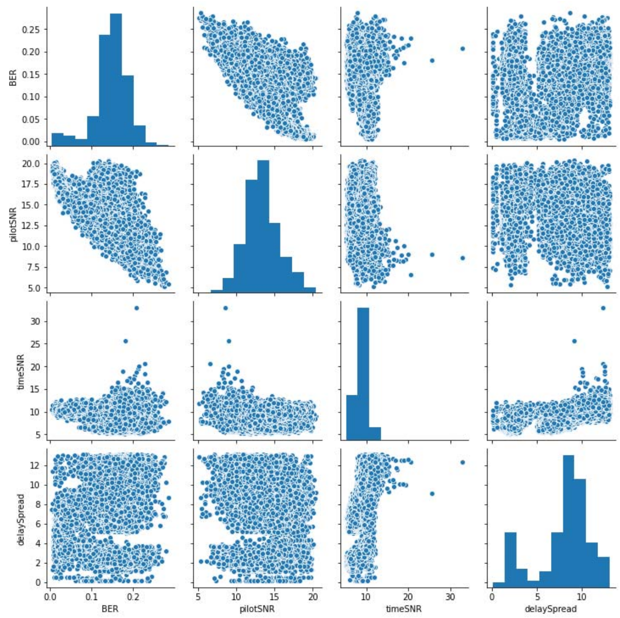
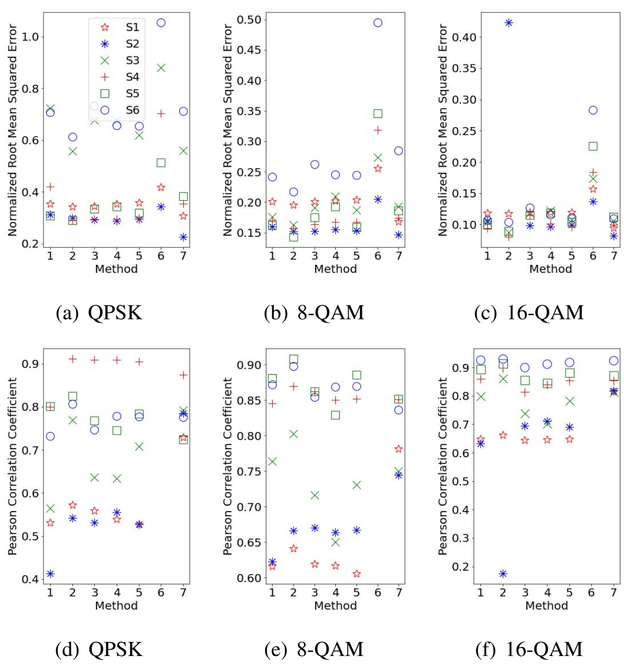


Fig. 6. SPACE08: An example of the crossplot at S3 and for 8-QAM. The SNRs are in dB; and the unit of the delay spread is milliseconds.

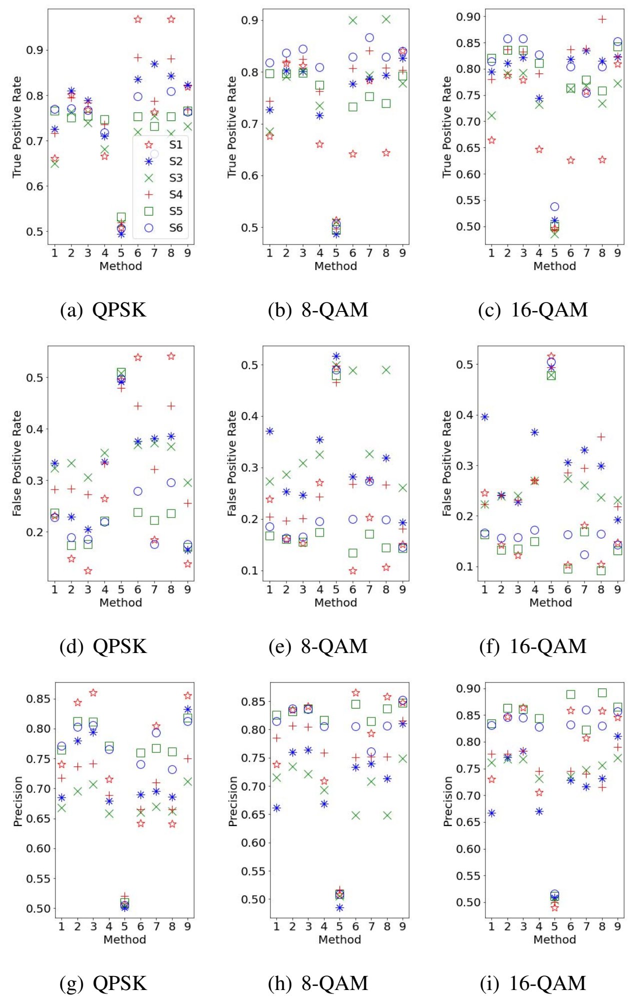


(a) QPSK (b) 8-QAM (c) 16-QAM

(d) QPSK (e) 8-QAM (f) 16-QAM

Fig. 7. SPACE08: Numerical prediction performance. Method 1: linear regression; method 2: RBF linear regression; method 3: neural net with 1 layer; method 4: neural net with 2 layers; method 5: KNN; method 6: ZeroR; and method 7: decision tree.

The performance of several classification methods are depicted in Fig. 8. Most of the classifiers perform well. The best consistent performance is seen on the random forest and the two neural networks. Sometimes the two-layer neural network outperforms the single-layer version or vice versa, but generally they perform similarly. The decision rule, the decision tree and the naive Bayesian network slightly outperform the logistic regression and KNN. The ZeroR method has the worst performance.



(a) QPSK (b) 8-QAM (c) 16-QAM

(d) QPSK (e) 8-QAM (f) 16-QAM

(g) QPSK (h) 8-QAM (i) 16-QAM

Fig. 8. SPACE08: Classification performance. Method 1: logistic regression; method 2: neural net with 1 layer; method 3: neural net with 2 layers; method 4: KNN; method 5: ZeroR; method 6: decision rule; method 7: decision tree; method 8: naive Bayesian network; and method 9: random forest.

VI. Conclusions

This work investigated supervised learning for performance prediction in UWA communications. After quantifying the transmitter, the channel and the receiver by numerical and categorical parameters, the performance prediction was cast individually as a numerical prediction problem and as a classification problem. Meaningful insights were obtained by examining supervised learning methods using field experimental data sets. The data processing results showed that fairly good numerical prediction or classification performance can be achieved. For the numerical prediction, discriminative models such as decision trees and neural networks performed better than the generative models like linear regression and linear regression with RBF. For the classification, discriminative models also achieved better performance than the generative models such as logistic regression and the naive Bayesian network. The KNN method performed generally well for both numerical predication and classification.

In the future, the concept of transfer learning will be explored. In this work, separate models were created for each individual data set. With transfer learning, it could be

Authorized licensed use limited to: Michigan Technological University. Downloaded on December 20,2022 at 17:00:29 UTC from IEEE Xplore. Restrictions apply.

beneficial to apply a model trained on one data set to an entirely different data set. The parameters such as the relative water depth, distance between the transmitter and the receiver, and the modulation type, could be included as variables into the modeling process.

REFERENCES

[1] C. Wang, Z.-H. Wang, W. Sun, and D. R. Fuhrmann, "Reinforcement learning-based adaptive transmission in time-varying underwater acoustic channels," *IEEE Access*, vol. 6, pp. 2541–2558, 2018.

[2] K. Pelekanakis and L. Cazzanti, "On adaptive modulation for low SNR underwater acoustic communications," in *Proc. of MTS/IEEE OCEANS Conf.*, Charleston, Oct 2018.

[3] Y. Wang, M. Martonosi, and L.-S. Peh, "Predicting link quality using supervised learning in wireless sensor networks," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 11, 2007.

[4] V. Kalaiarasu, H. Vishnu, A. Mahmood, and M. Chitre, "Predicting underwater acoustic network variability using machine learning techniques," in *Proc. of MTS/IEEE OCEANS Conf.*, Anchorage, Sept 2017.

[5] S. Jiang, "On securing underwater acoustic networks: A survey," *IEEE Communications Surveys & Tutorials*, vol. 21, pp. 729–752, 2018.

[6] W. Sun and Z.-H. Wang, "Online modeling and prediction of the large-scale temporal variation in underwater acoustic communication channels," *IEEE Access*, vol. 6, pp. 73 984–74 002, 2018.

[7] S. Zhou and Z.-H. Wang, *OFDM for Underwater Acoustic Communications*. Wiley, Jun. 2014.

[8] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning*. Springer Series in Statistics New York, 2001, vol. 1, no. 10.

[9] A. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2008, pp. 1177–1184.

[10] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees*. CRC press, 1984.

[11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-Learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[12] K. Pearson, "VII. Note on regression and inheritance in the case of two parents," *Proceedings of the Royal Society of London*, vol. 58, pp. 240–242, 1895.

[13] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[14] Z.-H. Wang, S. Zhou, G. B. Giannakis, C. R. Berger, and J. Huang, "Frequency-domain oversampling for zero-padded OFDM in underwater acoustic communications," *IEEE J. Ocean. Eng.*, vol. 37, pp. 14 – 24, Jan. 2012.

Authorized licensed use limited to: Michigan Technological University. Downloaded on December 20,2022 at 17:00:29 UTC from IEEE Xplore. Restrictions apply.