Data Based Sparse Learning for Underwater Sensor Array Networks

Shougui Cai
College of Information Science and Electronic Eng.
Zhejiang University
Hangzhou, China
shouguicai@zju.edu.cn

Wen Xu
College of Information Science and Electronic Eng.
Zhejiang University
Hangzhou, China
wxu@zju.edu.cn

ABSTRACT

Develop methods to effectively clean and reduce the amount of data at sensor level is very important in underwater networks, due to the limitation of bandwidth and power supply. In this paper, we look at this issue from the perspective of machine learning and get a sparse repesentation model by training the feedfoward neural networks. Results of range estimation on SWellEx96 experiment are compared to traditional MFP method. Simulation show that the neural network model have more ability to resist sound speed profile mismatch and can be fine-tuned to a information extractor.THe extractor reduces original data to small useful finite sets and make it more convenient to share and transfer information in underwater networks.

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability;

KEYWORDS

Underwater sensing,Maching Learning,Sparse Representation

ACM Reference format:
Shougui Cai and Wen Xu. 2017. Data Based Sparse Learning for Underwater Sensor Array Networks . In *Proceedings of 12th ACM International Conference on Underwater Networks and Systems, Halifax,NS,Canada, November 2017 (WUWNet'17),* 4 pages.
https://doi.org/10.475/123_4

1 INTRODUCTION

As the increasing number of manned platforms deploying receiver arrays or fixed/mobiled nodes in distributed underwater sensor networks,we are facing the scene of underwater big data.Due to the limitation of energy used,it is not realistically to process all the incoming data[3].Now,machine learning technology is more and more mature, we could try to introduce some smart into our system,to extract task-relevent information,discover new patterns,construct a useful representation.It is convenient for us to share and transfer information between different platforms once the data can be reduced to small useful finite sets.

WUWNet'17, November 2017, Halifax,NS,Canada

ACM ISBN 123-4567-24-567/08/06. . . $15.00
https://doi.org/10.475/123_4

2 MACHINE LEARNING BASED DATA SPARSE REPRESENTATION

In Underwater Networks,we will get a lot of data. What we really care about,is not the signal itself,but the information contained in the siganl.Meanwhile,the communication for underwater sensors is costly.So develop methods to represent the information present in metadata with a sparse format is very meaningful.

In this section,we discuss an effective scheme using machine learning based methods.

2.1 Neural Networks Models and Function Approximation

As we known,neural networks models can be viewed as a mathematical function f.Taking feedforward neural network(FNN) as an example,they define a mapping y $= f(x;\theta)$ between input x and output y by parameter θ,which needed to be learned by a rule.Feedforward networks are called networks because they are typically represented by composing together many different functions.We might have two function f^1, f^2 connected in a chain[1], to form $f(x) = f^2(f^1(x))$.

FNN extend linear models to represent nonlinear transformed format $\phi(x)$ of input x.The transform function ϕ can be thinked as providing a set of features describing x, or as providing a new representation for x.The key problem here is how to choose the mapping ϕ.

The strategy of machine learning is to learn ϕ.In a feedforward network,ϕ defines a hidden layer $h = \phi(x; w^{(1)})$,then the total model is $y = f(x;\theta, w) = hw^{(2)}$. obviously,We need to learn ϕ from a broad class of functions,and parameters w mapping $\phi(x)$ to the desired output.

In most cases,our parametric model defines a distribution $p(y|x;\theta)$.Simply, we can use the principle of maximum likelihood to learn parameters in model.

$$J(\theta) = -E_{x,y\sim p_{data}} \log p_{\text{mod } el}(y|x) \tag{1}$$

where the specific form of $p_{\text{mod } el}$ is defined by networks. As maximum likelihood is a consistent estimator, the model is capable of representing the training distribution.

2.2 Regularization for neural networks

There are two main kind of regularization strategys for neural networks,one is weight-level regularization,another is neuron-level regularization with activation penalty.

$$\tilde{J}(\theta; X, y) = J(\theta; X, y) + \alpha\Omega(\theta) + \beta\Omega(h) \tag{2}$$

where $\Omega(\theta)$ is parameter norm penalty, $\Omega(h)$ is penalty on the activations of the units,α,β are hyperparameters that weight the relative contribution of the norm penalty term.Weight decay term penalize the size of the model parameters,while,the activation penalty term encouraging their activations to be sparse.

2.3 Learn useful sparse representations from data

When a useful sparse representation of any given data is learned,each datum will then be encoded as a sparse code,thus we can use the least possible amount of resource to store or transfer the data.

$$\begin{aligned} &\tilde{J}(\theta) = -E_{x,y\sim p_{data}} \log p_{\text{mod el}}(y|x) + \lambda\|h\|_1 \\ &s.t. \quad \left\|W_i^{(1)}\right\|_2 \le C \quad \forall i = 1, \cdots, M \end{aligned} \tag{3}$$

In practical application, we not only want the representation to be sparse,but also want the model features to be sparse, the latter saves the storage and calculation on sensor nodes. Thus,we use $L1$ norm to promote sparse neurons activations, and constrain the norm of each column of the weight matrix to prevent any one hidden unit from having very large weights. In the equation, M is the number of neurons in hidden layer.

3 SIMULATION AND EXPERIMENTAL RESULTS

A notable recent example of using machine learning method in underwater acoustic is the application of nonlinear classification to source localization[2]. In this section,we implement a simple FNN with just one hidden layer to learn source range directly from observed acoustic data, and compare the performance of the classifier with the traditional matching field processing method (Bartlett) in terms of simulation data and experimental data, respectively.In addition, we briefly discuss the influence of sound speed profile(ssp) mismatch on the performance of FNN classifier and improve the tolerance of the classifier by training the model using data sampled under different ssp.

Simulation environment is the widely studied SWell96Ex test,conducting in a 216m deep shallow waveguide environment.The ship proceeded northward at a speed of 2.5 m/s.

The source ship has two sound source, a deep source (J-15) and a shallow source (J-13).In all the simulations, we used the shallow sound source, which was towed at a depth of about 9m and transmitted 9 frequencies between 109Hz and 385Hz.

3.1 Parameter Settings

In simulation part,acoustic data used to train and test the neural network is generated by kraken.Snapshot N_s is 10,number of vertical array elements L is 21,input layer neurons D of FNN is $L^2 \times N_{fre}$(number of frequency used),there is an input data preprocessing here, actual input data is the normalized sample covariance matrices(CSDM) of measured pressure at each frequency. The number of neurons in the output layer (number of classes) $K = 300$.We just give the same amount of dimensions the original data are encoded in to provide enough dimensions to learn over-complete features at first,which means neurons in hidden layer $M = D$. Specifically,the cost function now becomes

$$\begin{aligned} &\min \quad \tilde{J}(\theta) = -\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} \ln y_{nk} + \lambda\|h\|_1 \\ &s.t. \quad \left\|W_i^{(1)}\right\|_2 \le C \quad \forall i = 1, \cdots, M \end{aligned} \tag{4}$$

where t_{nk} and y_{nk} are predictive and real probability of sample data x belongs to class k separately.Moreover we will choose $C = 1$ here. For the sake of learning speed ,we use the $ReLU$ activation in hidden layer and use the $softmax$ function in the output layer because this case is a multiple class problem. The training set is 3000 samples of uniform sampling between 1.82-8.65km,test set is another 300 data samples sampling from the same range.The noise in the simulation is set to complex gaussian white noise.

Experimental data gets from SWell96Ex Event S5,we use the receive sound pressure of VLA to train our neural network.The array recorded a total of 75 min of data,In order to facilitate processing, we took 0-50min data as a training set.

Consistent with the simulation part,we divide the trajectory into 300 grids,25m each.We set 1 second as a snapshot and get 3000 sample covariance matrix (SCM),the sample covariance matrix is averaged at every two snapshots. At the time of training, we took 9/10(2700)of samples as training set and another 1/10(300) as test set.

3.2 The Effect of Sparse Constraint Training

The regularization degree on model affects the model error and average activation density of FNN's hidden layer. As the coefficient grows, the model error on training set and test set also grows, but slower on test set, which means the regularization helps improving the performance on test data in some sense. The model error is defined as the dissimilarity between true probability distribution and estimated probability distribution, thus the Kullback-Leibler(KL) divergence. In Fig.2, we use the cross entropy equivalently.

On the other hand, the regularization on neuron-level significantly reduces the average activation density, The order of magnitude drops from 10^3 to 10, and keep it stably. This phenomenon is a good news for us to train a sparse representation form data.

Compared to the case of training without sparse constraints, sparse constraint makes the weight coefficient in

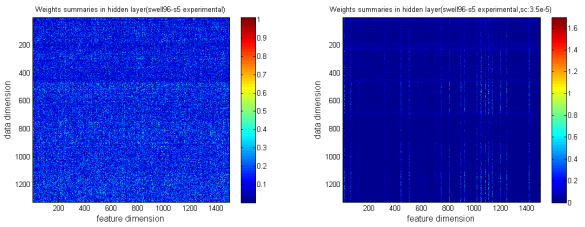

Figure 1: weights summaries in hidden layer(left:no constraint,right:with constraint).

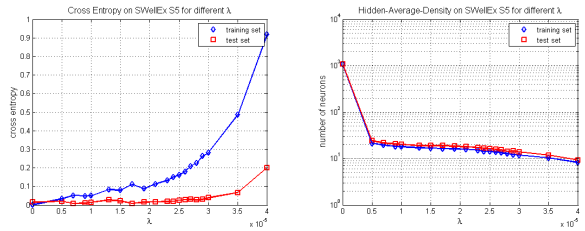

Figure 2: Model error and Average-activation-density for different λ.

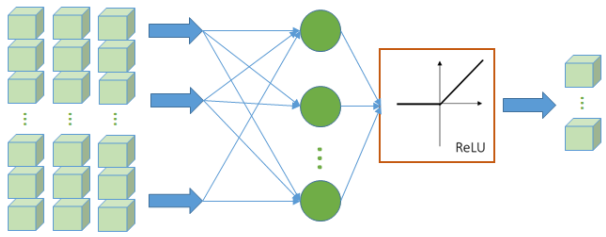

Figure 3: The learned sparse representation model.

the hidden layer show the group structure,either all zero, or basic is not zero. Take the case $\lambda = 3.5 \times 10^{-5}$ for example, the number of feature vectors reduced from 1500 to 140, as shown in Fig.1. In addition, it can be seen that the relative size of learned weights is related to the frequency,even at the same frequency, the the weight corresponding to real and imaginary parts is also different, there are bright and dark strip distribute along the data dimension.The data dimension is arranged according to the frequency relationship.

Apart from being beneficial to feature selection,sparse constraint reduce the activation rate of neurons in the hidden layer without reduce accuracy. In our example, when the coefficient is 3.5e-5, the average activation neuron number is 11,which is greatly reduced,the activation rate is only 0.7%. These mean the input 1323-dimension data can be represented by 140 feature vectors, averagely, each data sample can be represented by only 11 feature, as Fig.2 showns.To sum up, by using regularization strategy on neural networks level and weight-level of the FNN, we can get a sparse and low rank model, sparse means the transferred representation is sparse, low rank means the rank of learned weight matrix is low.

Table 1: Localization accuracy of FNN and MFP on SWell96Ex-S5 data

Methods	FNN	MCE	Bartlett 1	Bartlett 2
109Hz	89.3%	72.3%	37.7%	3.7%
232Hz	97%	91%	17.7%	4.3%
385Hz	99.7%	97.7%	14%	0.67%
109,232,385Hz	99%	99.7%	40.7%	7.7%

Table 2: Absolute mean error of FNN and MFP on SWell96Ex-S5 data(m)

Methods	FNN	MCE	Bartlett 1	Bartlett 2
109Hz	28.1	290.3	852.8	1219.5
232Hz	7.4	2.5	832.3	832.3
385Hz	0.08	0.58	1266.7	1756.3
109,232,385Hz	0.25	0.083	477.2	722.9

3.3 Comparison with traditional matching field processing method

As a reference,here we use Bartlett Processor to position the source position.There are two kinds of replica-field used in Bartlett Processor,one is calculated from model by kraken(noted as bartlett 2), another is from measurement data(noted as bartlett 1).

The accuracy and absolute error of methods under different frequency is summed in table 1 and table 2.As we can see,whether it is single frequency or combination frequency,the accuracy of FNN is always better than Bartlett,and not worse than direct data match(noted as MCE).FNN is helpful in positioning problem.

3.4 The influences of ssp mismatch on FNN classifier

In the MFP, the model accuracy is heavily affected by the mismatch.Fig.5 gives the FNN positioning results in different degrees change of sound speed profile. Here, snapshot is 10, SNR is 5dB. Comparing to optimized-ssp,the i905-ssp has only a very small change,within 0.5m/s at the same depth.The change in i906-ssp is much significant,whic can be seen from the shape in Fig.4.

Fig.5 plots the performance curves for FNN, Bartlett, MCE by 1000 times Monte Carlo simulations.When the change in ssp is relatively small, FNN positioning best,MCE second and Bartlett worst.When the shape of ssp change,the accuracy order unchanged,but the absolute error of FNN becomes bigger than MCE.

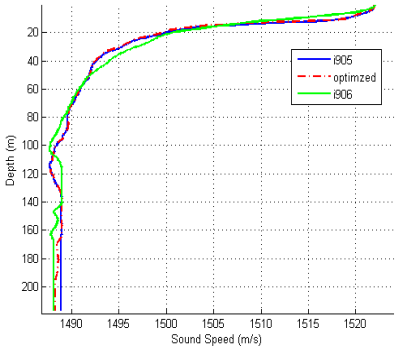

Figure 4: plot of sound speed profile.

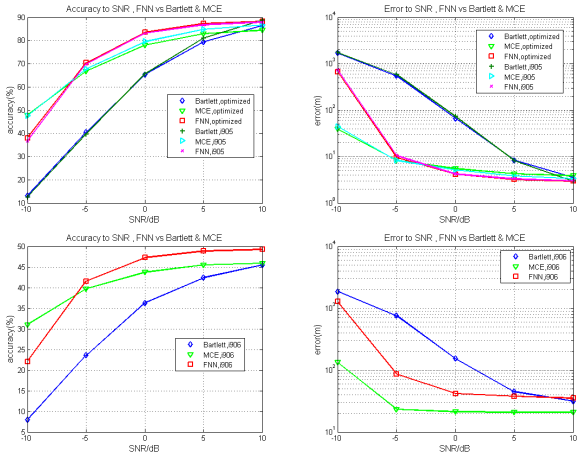

Figure 5: FNN positioning performance curve(frequency:109,232,385Hz).

3.5 Increase model tolerance by mixed data training

The simulation results show that train the model by data collected from different ssp can significantly improve the tolerance of the classifier,which means FNN can learn weigths over a set of changing ssp.This is an interesting and useful discovery.

In Fig.6, i906* is little changed from i906,for the sake of testing.Although the accuracy for i905 has a little glissade compared with single data training,the performance for i906 imporved.In general,the trained FNN classifier works well on both two different shape ssp.

4 CONCLUSIONS

From the veiw of information theory, the objective of signal processing is to exact specific task-relevant information from data. How to exact the relevant information from measured data is an important issue.Machine learning may be able to provide ideas for this issue, because of its strong ability to learn and characterize. In this paper,we develop a method

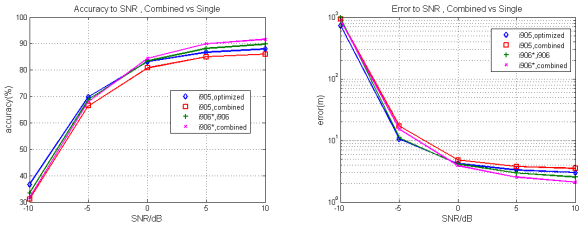

Figure 6: FNN positioning performance curve.

to learn sparse representation from data based on neural networks.At present, we only discuss the situation of a single linear array, the situation for multi-array will be dicussed in future. Verification on SWell96Ex experimental data shows this method can help select feature vectors and extract task-related information contained in the siganl. The pretained sparse repesentation model can help the underwater networks use the least possible amount of resource to store or transfer the data. Compared to the traditional MFP method, the neural network trained model have more ability to resist sound speed profile mismatch and can fine-tune to a feature extractor conveniently.

ACKNOWLEDGMENTS

The work is supported by ...

REFERENCES

[1] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep learning*. MIT press.
[2] Haiqiang Niu, Peter Gerstoft, and Emma Reeves. 2017. Source localization in an ocean waveguide using supervised machine learning. *arXiv preprint arXiv:1701.08431* (2017).
[3] T. C Yang. 2015. Issues in Underwater Acoustic Signal Processing in the Era of Big Data. (2015).