Radio Signal Detection Using Machine-Learning Approach

Ivan V. Malygin
Ural Federal University
Yekaterinburg, Russia
pit_pit2@mail.ru

Sergey A. Belkov
Ural Federal University
Yekaterinburg, Russia
buf2@mail.ru

Mihail R. Usvyatsov
Moscow Institute of Physics and Technology, Skolcovo Institute of Science and Technology
m.usvyatsov@gmail.com

Alexander D. Tarasov
Ural Federal University
Yekaterinburg, Russia
alex@chrns.com

Abstract

Typically, noise-like signals utilize correlators and matched filters for the purpose of signal detection. Their target is to maximize noise rejection and maximize signal efficiency. We propose the use of a neural network approach for the problem of signal detection.

1 Introduction

Nowadays, correlators and matched filters are used for noise-like signal detection and processing. The traditional methods of signal detection utilize correlation function thresholding. Despite the theoretical justification and practical results, the problem of low-amplitude signal detection over an extremely noisy environment is still relevant.

In Section II, the authors will demonstrate why the problem of signal detection over a noisy environment arises. Next, a formal problem statement will be done in Section III. The authors will present an experimental stand schema in Section V, which will help analyze the results obtained during the experiments in Section VI. Conclusions and open problems will be provided in Section VII.

2 Digitization

Electronic systems transmit information in signal form from a transmitter to a receiver via environment. Modern electronic devices utilize digital signal representation ($s \in \{0,1\}$) due to its relatively low cost, complexity, and the high availability of the hardware necessary for their processing in contrast to analog devices. Furthermore, digital signal processing guarantees a higher signal/noise ratio [2].

A signal can be transmitted through an environment only in continuous form, independent of the transmitter and receiver work mode. Digital signals could be obtained from continuous ones through a process of digitization that could be mathematically described as: $D : [0,1] \to \{0, 1\}$. The process of binary digitization can be described as follows:

$$D(x) = \begin{cases} 1, f(x) > \tau \\ 0, otherwise \end{cases}$$

For fixed τ. Digitization isn't a lossless process, as it is a transformation from a continuous space to a discrete one.

3 Problem formulation

Let the transmitter transmit a signal s_1. Next, the signal is transformed to analog form $s_2 = D^{-1}(s_1)$. Going through the environment, signal s_2 could be distorted, and so $D(s_2) \neq s_1$.

It is clear that 'the degree of similarity' $D(s_2)$ and s_1 are defined by a system noise stability property, however it is not a formal definition. Both signals are digital sequences and they could be compared with such metrics as: Hamming distance [3], Wasserstein distance [4], a maximum value of autocorrelation function [5], etc. A comparison of these metrics and the development of an optimization technique to compare $D(s_2)$ and s_1 are the subject of this work.

A similarity metric is required to determine whether a received signal is similar enough to the transmitted one (reference) or they are too different and one cannot argue that the expected signal was received.

Machine learning formulation of the problem is a binary classification problem. Formally, let exist classifier $F : S$ is ¿ C, where S is a feature space of classifier input $v \in S$, C is a classes space. In this case, there are two possible classes:

1) $D(s_2)$ and s_1 are close in terms of the metric (equivalent);
2) $D(s_2)$ and s_1 are very much different ($D(s_2)$ - noise).

It is important to show how $D(s_2)$ and s_1 could be represented in the feature space. Let us consider a case when classifier input is only signal $D(s_2)$ and choose its particular realization v. Then a bit size of v sequence is a dimension of feature space. Next, signal v could be represented as a ddimensional vector in feature space, where v_i is i-component of vector v in space S. Let's consider an example: the signal s_1 is defined by a protocol that has a 3 bit length, with the bits equal to (101). C space can be divided into two subspaces: C_1, which consists of signals equivalent to a defined protocol signal; C_2, which consists of noise signals. The task of classifier F is to build such a hyperplane in space S that divides C_1 and C_2 in the 'best' way (in terms of the selected classifier quality metric).

An alternative way to represent classifier input vin that consists of a d bit:

$$D(x) = \begin{cases} v_{in}^i, if\ D(s_2)^i = s_1^i \\ 0, otherwise \end{cases}$$

One can start from an initial approximation in order to build the desired hyperplane in the space S. Next, the quality metric must be defined. And also training and validation datasets. In this case, datasets consist of signals belonging to class C_1 and class C_2. For every example in a dataset, ground truth should be known. In the best case, classes should be balanced, otherwise classifier learning won't lead to the optimum of the selected quality metrics [7]. After redefining our task as a machine learning problem, it becomes possible to apply the stochastic gradient descent algorithm [6] to teach the neural network by means of a backpropagation technique as a method of classifier optimization by its quality metric. This approach is preferable over other methods that fix the features that are considered. That is caused by the fact that the neural network is able to learn features that contain information from the data that leads to a quality function optimization.

4 Neural Network

In order to describe neural network mechanics, it is necessary to describe the behavior of a single neuron.

A diagram of a neuron's behavior is depicted on Fig. 1. If neuron input x is m-dimensional vector, then a neuron could be mathematically described as $f(x) = g(xW^T + b)$, where W is a weights vector that is multiplied by an input vector, and b is a bias term. Obviously, the argument of function g is a number. Function g is a nonlinear part of a neuron.

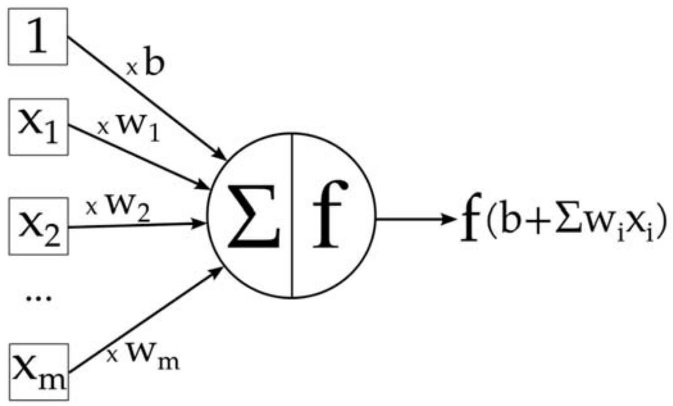

Figure 1: Neuron model

A conventional neural network consist of neurons organized in a structure. All computations for such neural networks can be written in a matrix form that allows us to compute them effectively.

According to the approximation theorem, a neural network is able to approximate any continuous function [8]. This property is actually used when we approximate a classifier with the neural networks.

During the process of learning, the output of the neural network for a given input (dataset) is computed and these answers are compared with the ground truth. Incorrect answers are then used for the stochastic gradient descent algorithm, which updates the neural network weights using this information.

Thus, the neural network can learn functions by examples, using a quality functional (loss function).

5 Experiment

The aim of this work is to compare a correlator and a neural network. For a comparison of these models, it is necessary for a set of signals, for which marked dataset (signal, noise) has been passed through the neural network and the correlator. The difficulty in this task is that, when building such a dataset, there is not an objective label for each example. For example, the signal was transmitted, but the noise level was so high that the signal did not reach the receptor. On the other hand, we can make reliable conclusions about the presence of a useful signal in what was received, analyzing the level of noise in the propagation medium.

The easiest way to obtain such data is to add artificial data to the signal and to use some rule to assign ground truth. However, for this particular setup a neural network can only learn the rule we have used to assign the ground truth.

A better way is to use real data that has gone through a real environment. After the data is received, the neural network will have to learn the noise impact on the signal. Every bit of information is encoded with m-sequence 1023 bit length.

A developed experiment schema is depicted on Fig. 2. This schema will help make a pre-conclusion about the applicability of neural network usage for a signal classification problem. The transmitter and receiver in this schema are a single microcontroller. This approach makes it possible to synchronize the transmitter and the receiver.

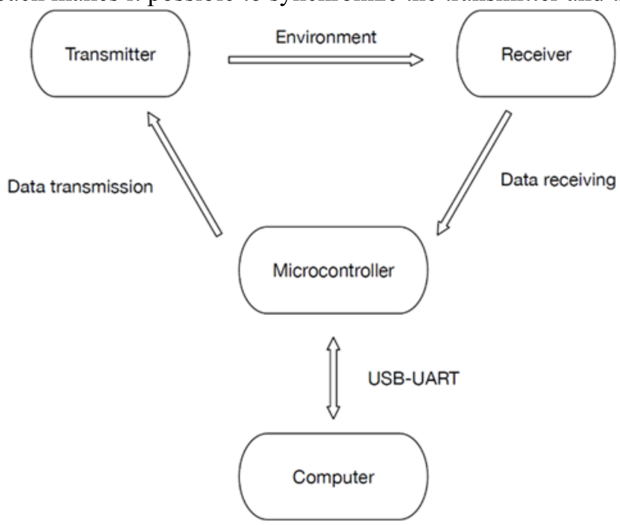

Figure 2: Experiment schema

The following setups were used during dataset preparation:

1) antennas are on the distance of 5 cm;
2) the same as previous, but with presence of background noise;
3) antennas of transmitter and receiver are in a metal jar.

A noise level analyzer for the environment is not yet realized and its design is a crucial point of further research.

6 Result

Using the proposed schema, experimental data consisting of these sequences sets was obtained: transmitted, received and ground truth (obtained using synchronization). Noise in the channel that was received during the absence of transmitted data was used as negative examples.

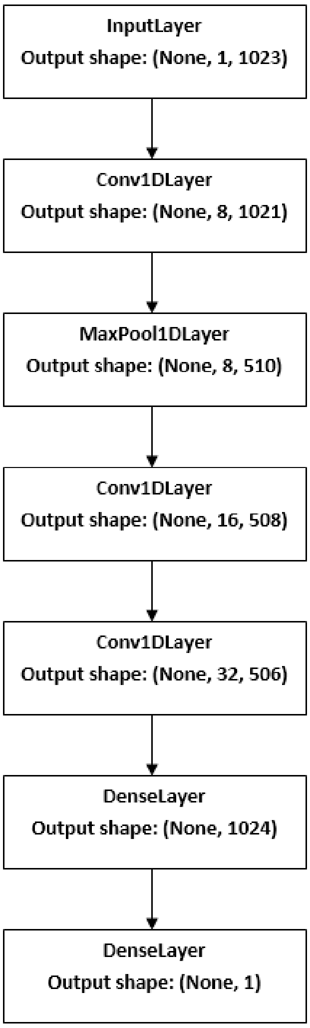

Figure 3: Network architecture

The neural network (architecture is depicted on Fig. 3) achieved 99 % accuracy on a test dataset for different transmission setups using binary cross entropy as a loss function [9]. The selection of this architecture is governed by the locality property of bit influence (close bits are more correlated than those further away). In those instances, where the correlator response was negative (it was decided that the signals were not correlated), the neural network was able to output a positive answer that was correct according to ground truth. However, due to the fact that the dataset was obtained in the absence of noise level analysis in the environment, the neural network was trained on a large dataset of false positive examples. That led to an incorrect feature extraction by the neural network and a higher amount of false positives in the neural network's answers.

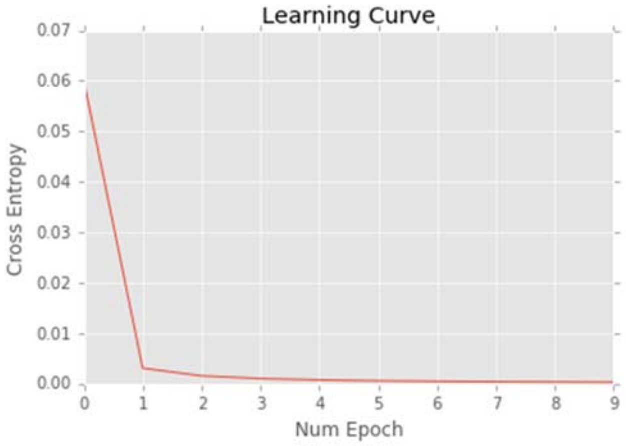

Figure 4: Learning curve

According to Fig. 4, only a few iterations of the gradient method are enough to reach the minimum (a value close to the minimum, to be precise) of the objective function.

7 Conclusion

In this work we demonstrated how to build an experiment to obtain data that could be used for a neural network application for the problem of radio signal detection. According to the obtained results, the correlator could be approximated with the neural network and we propose that, in general, it can be replaced by the neural network. The problem of neural network application is the lack of annotated data. Once ground truth is obtained, it will be possible to adjust the neural network to a specific environment and noise.

References

1. Barker H. Group synchronizing of binary digital sequences //Communication theory. Butterworth, London, 1953. pp. 273-287.
2. Harris D., Harris S. Digital design and computer architecture. – Elsevier, 2012.
3. Forney G. Generalized minimum distance decoding //IEEE Transactions on Information Theory. 1966. pp. 125-131.
4. Ruschendorf L. The Wasserstein distance and approximation theorems //Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete, 1985. pp. 117-129.
5. Welch L. Lower bounds on the maximum cross correlation of signals (Corresp.) //IEEE Transactions on Information theory, 1974. pp. 397-399.
6. Amari S. Backpropagation and stochastic gradient descent method //Neurocomputing, 1993. pp. 185-196.
7. K. Vorontsov Machine learning (course lectures) //electronic version: http://bit.ly/ml course, 2009.
8. Chen T., Chen H. Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems //IEEE Transactions on Neural Networks, 1995. pp. 911-917.
9. Shore J., Johnson R. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy //IEEE Transactions on information theory, 1980. pp. 26-37.