URSI GASS 2020, Rome, Italy, 29 August - 5 September 2020

Radio Frequency Interference Detection in Microwave Radiometry Using Support Vector Machines

Imara Mohamed Nazar*, Mustafa Aksoy
Electrical and Computer Engineering Department
University at Albany, SUNY
Albany, NY, USA
Emails: {imohamednazar, maksoy}@albany.edu

Abstract

In this paper, a machine learning based approach for radio frequency interference (RFI) detection in microwave radiometry is discussed. Features of RFI-free and RFI-contaminated radiometer measurements are different in nature; thus, a classification framework utilizing the support vector machine (SVM) is proposed to detect RFI contamination. The framework has been evaluated with a set of simulations where RFI-free radiometer measurements are modeled as Gaussian noise, and RFI contamination is simulated by injecting pulsed sinusoidal signals with variable amplitude and duty cycle into the radiometer measurements. The results of these simulations have demonstrated the ability of the proposed approach to detect RFI accurately even in low interference to noise ration (INR) and very short duty cycle (DC) cases.

1 Introduction

Although the use of radio frequency spectrum is regulated in a way that certain frequency bands are solely allocated for microwave radiometry applications, the amount of RFI observed in microwave radiometer measurements is significantly increasing over time [1]. Thus, it is important to design interference detection techniques while keeping track of frequency allocations. So far detection algorithms in spatial, temporal, spectral, statistical and polarimetric domains have been developed and implemented in radiometer systems [2]. Performance of some of these detection algorithms against pulsed sinusoidal interference have also been discussed in the literature [3, 4, 5, 6]. However, the variability of spectral, temporal, spatial, and statistical properties of RFI signals require multi-domain approaches for proper RFI detection and mitigation[7]. National Aeronautics and Space Administrations (NASA) Soil Moisture Active Passive (SMAP) mission operates one of the first radiometers which implements such multi-domain RFI detection procedure by combining the outputs of several algorithms running simultaneously in the temporal, spectral, statistical, and polarimetric domains with a logical OR operator [8]. However, SMAP products have also been reported to be susceptible to RFI, especially when the contamination is wideband and continuous [9]. Thus, new approaches, possibly in new domains, are required for RFI detection and mitigation in microwave radiometry.

In this paper, we introduce a support vector machine classifier to address the challenging problem of RFI detection in the feature domain through a set of simulations with pulsed sinusoidal interference. Sections 2-5 describes the simulations and operation of the classifier. Then, in Sections 6 and 7, the results are summarized, and conclusions and future work are discussed.

2 Simulation Setup

In this study, RFI is considered to be a pulsed sinusoidal signal with 1400 Hz frequency, 1 V amplitude, 0.5 s pulse repetition interval, and variable DC. To simulate RFI-contaminated radiometer measurements, this RFI is injected into a Gaussian noise with variable standard deviation which represent radiometer measurements. In order to analyze how the performance of the proposed algorithm changes with DC and INR, DC of the interference signal is varied from 1% to 100% in increments of 1%, and INR is increased from -20 dB to 10 dB in steps of 1 dB. Notice that that since the amplitude of the interference signal is kept constant at 1 V, INR is changed by altering the standard deviation of the Gaussian noise. Furthermore, it is assumed that the radiometer measurements are sampled at a 20 kSPS rate and the integration time is 0.5 s.

RFI-free and RFI-contaminated radiometer measurements are generated as explained above, and parameters that characterize them which are known as 'features', are extracted from each 0.5 s radiometer integration periods at multiple instances. Features are namely, the root mean square power, mean, variance, ratio between the mean and the variance, skewness, kurtosis, maximum amplitude, minimum amplitude and the number of the peaks which are the positive voltage samples with zero gradients.

3 Support Vector Machine

We consider a set S of integration periods. Each period $s \in S$ is denoted by d–dimensional feature vector $\mathbf{x} = \{z_1, \ldots, z_d\}$ called support vector where d is the number of features. Furthermore, each radiometer integration period s belongs to one of N number of classes. In this study, N is two, namely, RFI-free and RFI-contaminated measurements. A set of class labels $y = \{-1, 1\}$ is defined to denote these RFI-free and RFI-contaminated classes, respectively. SVM [10] denotes the class of hyper–planes of $\langle \mathbf{w}, \mathbf{x} \rangle + b = 0$, where $\mathbf{w}, \mathbf{x} \in \mathbb{R}^d$ and $b \in \mathbb{R}$. It is possible to prove that the optimal hyper–plane is the one with maximal margin of separation between the two classes. The optimization problem for maximal margin can be formulated as follows:

$$\begin{aligned} &\min \langle \mathbf{w}, \mathbf{w} \rangle, \\ &\text{s.t } y_s[\langle \mathbf{w}, \mathbf{x}_s \rangle + b] \geq 1. \end{aligned} \tag{1}$$

where, $\mathbf{x}_s$ and y_s denote the feature vector and the class label of integration period s, respectively. Consequently, the final decision function is given by:

$$f(x) = sgn(\langle \mathbf{w}, \mathbf{x} \rangle + b). \tag{2}$$

4 Feature Selection

In order to reduce the computational cost, a smaller subset of aforementioned features can be selected without undermining the detection performance. In this work, a similarity–based filter type feature selection method, namely, Fisher score is used [11]. Fisher score is easy to implement and computationally efficient. This method selects features considering that the feature values of integration periods belong to the same class are similar whereas the feature values of integration periods from different classes are dissimilar. Fisher score for each feature $z_i, i = \{1, \ldots, d\}$ is defined as follows:

$$FS_{zi} = \frac{\sum_{j=1}^{N} n_j (\mu_{ij} - \mu_i)^2}{\sum_{j=1}^{c} n_j \sigma_{ij}^2}, \tag{3}$$

where N, n_j, μ_i, μ_{ij} and σ_{ij}^2 denote the number of classes, number of samples in class j, mean value of feature z_i, mean value of feature z_i for samples in class j, and variance of feature z_i for samples in class j, respectively. Larger Fisher score value for a feature implies greater distinctiveness.

5 RFI Detection

The process of the proposed RFI detection approach is shown in Fig 1. Initially, the features listed in section 2 are extracted from the radiometer measurements within each integration period and stored in a dataset. Then, this data is divided into train and test datasets. Train dataset is used to train the SVM model, after applying the feature selection process described in section 4. Finally, the corresponding features are taken from test data, and then imported into the trained SVM model to label the integration periods as RFI-free or RFI-contaminated.

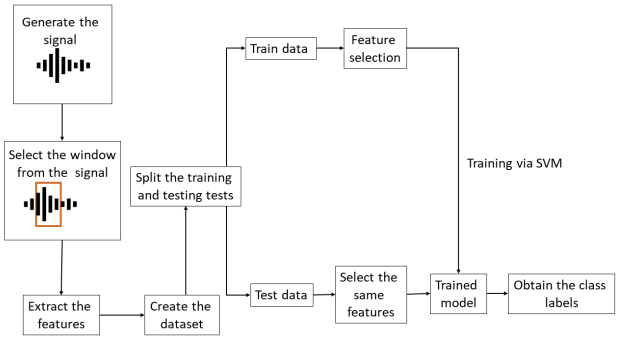

Figure 1. Interference detection process.

6 Results

Three thousand one hundred datasets are created to analyze the effect of INR and DC on RFI detection performance. Each dataset, associated with a pair of INR and DC values, consists of 398 radiometer integration periods which belong to either RFI-free or RFI-contaminated classes. In order to avoid over fitting, fivefold cross–validation was implemented in which the dataset was divided into five folds and each fold was tested by the model trained by remaining folds. Thus, in each fold, the SVM model was trained on 318 integration periods and tested on 90 integration periods. In each period, relevant features are selected based on the algorithm discussed in section 4. When multiple features have the same Fisher score, only one them is selected for RFI detection. Also, features that produce indefinite Fisher score values are discarded.

The RFI detection performance is evaluated based on the accuracy, precision and recall parameters which are defined as follows:

$$\begin{aligned} Accuracy &= \frac{TP+TN}{TP+FP+TN+FN}, \\ Precision &= \frac{TP}{TP+FP}, \\ Recall &= \frac{TP}{TP+FN}, \end{aligned}$$

where TP is the number true positives, i.e., number of RFI-contaminated integration periods identified as RFI-contaminated, TN is the number of true negatives, i.e., number of RFI-free integration periods identified as RFI-free, FP is the number of false positives, i.e., number of RFI-free integration periods identified as RFI-contaminated, and FN is the number of false negatives, i.e., number of RFI-contaminated integration periods identified as RFI-free.

Fig 2 illustrates the accuracy, precision, and recall of the proposed RFI detection method as a function of INR and DC. It can be observed that, in general, the accuracy, pre-

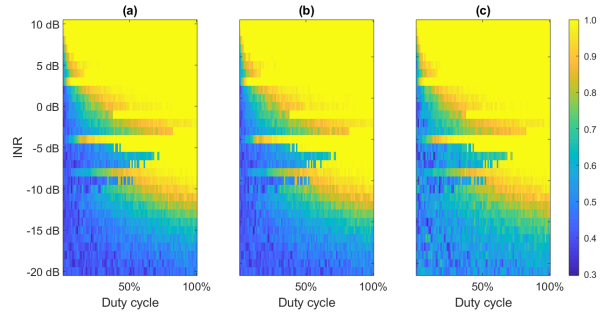

Figure 2. Accuracy (a), precision (b) and recall (c) as a function of RFI strength and duty cycle.

cision, and recall values increase with increasing SNR and DC. More than 95% of accuracy is achieved even very low, e.g, -10 dB, INR rates provided that DC is larger than 50%. This indicates the performance of the proposed algorithm against continuous RFI contamination. For lower DC, i.e., short duration RFI contamination cases, the accuracy diminishes, but still an acceptable detection performance can be reached if INR is larger than 0 dB. Slightly higher recall values, i.e., the fraction of detected RFI-contaminated periods out of all RFI-contaminated integration periods, especially for low INR cases, failure to detect RFI contamination happens less often. Finally, high precision values, i.e., the ratio of the number of RFI-contaminated integration periods to the number of periods labeled as RFI-contaminated, imply low false alarm rates for the proposed approach when accuracy and recall rates are high. Thus, the trade-off between RFI detection and false alarm rates are found to be minimal.

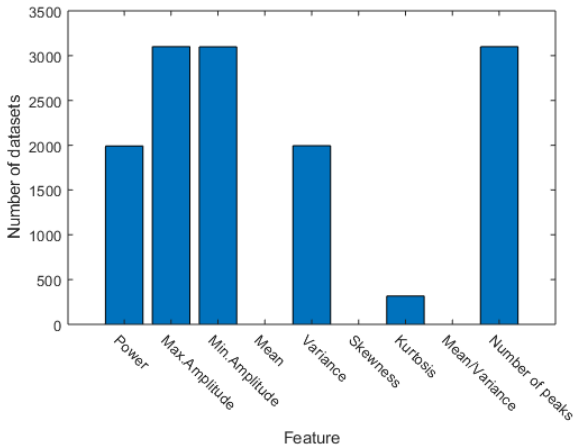

Figure 3. Frequency of features used in datasets.

Fig 3 shows the number of datasets that each feature is used for RFI detection in this study. As seen from the figure, the minimum amplitude, maximum amplitude and the number of the peaks are used more frequently than the other features. It is also observed that the root mean square power and the variance are used mostly in extreme (too low or too high) INR cases. Furthermore, the kurtosis is utilized in high INR cases.

7 Conclusions and Future Work

In this paper, an SVM classifier is introduced as a new approach for RFI detection in microwave radiometry through a set of simulations. Results illustrate the ability of this method in detecting pulsed sinusoidal interference. Furthermore, the distinctive features for RFI detection are demonstrated for the proposed classifier. In the future, we plan to compare this approach with traditional RFI detection algorithms.

References

[1] D. McKague, J. Puckett, and C. Ruf, "Characterization of k-band radio frequency interference from amsr-e, windsat and ssm/i," August 2010, pp. 2492 – 2494.

[2] S. Misra and P. De Matthaeis, "Passive remote sensing and radio frequency interference (rfi): An overview of spectrum allocations and rfi management algorithms [technical committees]," *IEEE Geoscience and Remote Sensing Magazine*, **2**, no. 2, pp. 68–73, June 2014.

[3] J. T. Johnson and L. C. Potter, "A study of algorithms for detecting pulsed sinusoidal interference in microwave radiometry," in *IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium*, July 2008, vol. **2**, pp. II–161–II–164.

[4] R. D. De Roo, S. Misra, and C. S. Ruf, "Sensitivity of the kurtosis statistic as a detector of pulsed sinusoidal rfi," *IEEE Transactions on Geoscience and Remote Sensing*, **45**, no. 7, pp. 1938–1946, July 2007.

[5] B. Guner, M. T. Frankford, and J. T. Johnson, "A study of the shapiro–wilk test for the detection of pulsed sinusoidal radio frequency interference," *IEEE Transactions on Geoscience and Remote Sensing*, **47**, no. 6, pp. 1745–1751, June 2009.

[6] R. D. De Roo and S. Misra, "A moment ratio rfi detection algorithm that can detect pulsed sinusoids of any duty cycle," *IEEE Geoscience and Remote Sensing Letters*, **7**, no. 3, pp. 606–610, July 2010.

[7] M. Aksoy, J. T. Johnson, S. Misra, A. Colliander, and I. O'Dwyer, "L-band radio-frequency interference observations during the smap validation experiment 2012," *IEEE Transactions on Geoscience and Remote Sensing*, 54, no. 3, pp. 1323–1335, March 2016.

[8] J. R. Piepmeier, J. T. Johnson, P. N. Mohammed, D. Bradley, C. Ruf, M. Aksoy, R. Garcia, D. Hudson, L. Miles, and M. Wong, "Radio-frequency interference mitigation for the soil moisture active passive microwave radiometer," *IEEE Transactions on Geoscience and Remote Sensing*, 52, no. 1, pp. 761–775, Jan 2014.

[9] P. N. Mohammed, M. Aksoy, J. R. Piepmeier, J. T. Johnson, and A. Bringer, "Smap l-band microwave radiometer: Rfi mitigation prelaunch analysis and first year on-orbit observations," *IEEE Transactions on Geoscience and Remote Sensing*, 54, no. 10, pp. 6035–6047, Oct 2016.

[10] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, **20**, no. 3, pp. 273–297, Sep 1995.

[11] J.Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature selection: A data perspective," *CoRR*, abs/1601.07996, 2016.