Template-Based Transient RFI Feature Recognition Algorithm: A Preliminary Study (Postprint)

Authors: Zhang Yazhou, Zhang Hailong, Zhang Meng, Wang Jie, Ye Xinchen, Wang Wanqiong, Li Jia, Du Xu

Date: 2021-09-29T00:00:00+00:00

Abstract

Transient radio frequency interference (RFI) signals represent the most challenging category of interference to identify and eliminate. To address the issue of transient RFI encountered during radio astronomy observations, this paper proposes a feature extraction methodology based on RFI signal peaks. This approach is utilized to generate feature templates for RFI signals, and a feature recognition algorithm is designed based on the Dynamic Time Warping (DTW) algorithm combined with a scoring strategy, thereby implementing RFI signal identification and classification capabilities. Cross-validation comparisons of the feature templates were conducted using RFI data provided by the ARRL official website. Test results demonstrate high similarity between RFI feature templates and RFI signals of the same category, indicating that the proposed algorithm can effectively recognize and classify RFI. This algorithm offers a novel approach for feature extraction and recognition of transient RFI.

Full Text

Preliminary Study of Transient RFI Feature Recognition Algorithm Based on Template

Zhang Yazhou[12], Zhang Hailong[1234], Zhang Meng[12], Wang Jie[14], Ye Xinchen[14], Wang Wanqiong[1], Li Jia[1], Du Xu[12]

[1] Xinjiang Astronomical Observatory, Chinese Academy of Sciences, Urumqi, Xinjiang 830011, China
[2] University of Chinese Academy of Sciences, Beijing 100049, China
[3] Key Laboratory of Radio Astronomy, Chinese Academy of Sciences, Nanjing, Jiangsu 210008, China
[4] National Astronomical Data Center, Beijing 100101, China

Abstract: Transient radio frequency interference (RFI) signals are among the most challenging types of interference to identify and eliminate. To address the problem of transient RFI encountered during radio astronomical observations, this paper proposes a feature extraction method based on RFI signal peaks. This method generates characteristic templates for RFI signals, and a feature recognition algorithm is designed based on the Dynamic Time Warping (DTW) algorithm and a scoring strategy to achieve RFI signal identification and classification. Cross-validation of the feature templates was completed using RFI data provided by the ARRL website. Test results demonstrate that the RFI feature templates exhibit high similarity with RFI signals of the same category, indicating that the algorithm can effectively identify and classify RFI. This approach provides a novel solution for feature extraction and recognition of transient RFI.

Keywords: RFI; Periodic computation; Feature extraction; Feature recognition

0 Introduction

With advances in radio astronomy technology, radio telescopes are evolving toward larger apertures and array configurations. Observational facilities such as China' s Five-hundred-meter Aperture Spherical radio Telescope (FAST) [1], the Low-Frequency Array (LOFAR) in the Netherlands [2], and the Very Large Array (VLA) in New Mexico, USA [3], possess high sensitivity and resolution capabilities. While these instruments enable the detection of extremely weak celestial signals, they also receive interference signals. Radio frequency interference (RFI) signals, which are artificially generated and often several orders of magnitude stronger than faint astronomical signals, must be effectively suppressed during radio astronomical data processing.

RFI originates from diverse sources. Literature [4] categorizes RFI into external and internal types. External RFI refers to wireless interference generated outside the observatory site, commonly including artificial satellites [5] (such as BeiDou and GPS navigation satellites), aircraft, nearby base station signals, and broadcast signals. Internal RFI refers to interference produced by electronic devices used within the observatory site, such as various digital terminal equipment, computers, microwave ovens, coffee machines, fluorescent lights, and other electronic instruments.

RFI mitigation has been a persistent research focus throughout the development of radio astronomy. Literature [6] notes that astronomical signals are typically broadband and often vary smoothly over time, whereas most RFI exhibits distinct differences compared to astronomical signals. For RFI mitigation, literature [7] proposed the CUSUM method, which is simple and efficient but relies on single sampling points and is sensitive to noise. To overcome this limitation, literature [8] introduced the SumThreshold algorithm based on combined threshold representations of adjacent sampling points. Literature [9] utilized singular value decomposition to address broadband RFI issues in the Giant Metrewave

Radio Telescope (GMRT). These algorithms are essentially threshold-based and perform well when processing RFI signals with amplitudes far exceeding astronomical signals. However, for RFI signals with characteristics similar to astronomical signals, threshold-based methods may mistakenly remove astronomical signals, leading to information loss and affecting the accuracy and effectiveness of astronomical research.

Transient RFI refers to radio frequency signals emitted during the normal operation of certain devices or systems. Relays, electric motors, and switching power supplies all generate transient interference. Since transient interference signals are broadband, pulsed, and have very short durations, their sources are difficult to identify. Literature [10] proposed a dictionary-based transient RFI classification algorithm that uses a Hidden Markov Model to identify transient RFI as subsequences and can extract transient RFI sequences from observational data. This paper proposes a feature extraction and recognition method specifically for transient RFI, aiming to achieve effective feature extraction and identification for particular types of transient RFI.

1.1 Signal Period Computation

Observing most RFI signals on the time scale reveals clear periodic fluctuations. To analyze the characteristic information of RFI signals, windowing processing is required, where the window size is closely related to the signal' s periodic information. Computing the signal period is therefore an important prerequisite for analyzing its characteristic information.

Traditional signal period computation algorithms include the cepstrum method, autocorrelation function method, average magnitude difference function method, and wavelet transform method [11]. The autocorrelation function method is a highly effective algorithm for period estimation of low signal-to-noise ratio signals [12]. Based on the autocorrelation function method, this paper designs an autocorrelation convolution sliding algorithm to compute the period of RFI signals. The algorithm proceeds as follows: (1) Take the absolute value of the amplitude of an RFI signal segment of length N. (2) Extract a signal segment of length M from the RFI signal as the sliding window signal. (3) Perform convolution operations between the sliding window and the RFI signal starting from the initial position, sliding backward by d sampling points each time. (4) Repeat step (3) until the sliding window reaches the end of the RFI signal. (5) Calculate the signal period using the peak information from the convolution results.

The sliding window size M should be smaller than the RFI signal length N. In Figure 1, the sliding window size M is set to 181. To compute the period accurately, the sliding step size d is set to 1. When the sliding window reaches the end of the RFI signal, zero-padding is required if the remaining signal length is less than the sliding window signal length. Figure 1 shows the period analysis of the k7ing-3545khz signal (an mp3 format RFI signal from the ARRL website), selecting approximately 0.29 seconds of signal from 0.65s to 0.94s as shown in Figure 1(a).

1.2 RFI Signal Feature Extraction

Peaks are important characteristic information in RFI signals that can capture the variation details of signal fluctuations. The RFI signal feature extraction process consists of the following steps: (1) Select an RFI signal window. (2) Take the absolute value of the amplitude of the window signal. (3) Obtain the peak information of the signal window. (4) Preprocess the discrete peak information to smooth it. (5) Use the processed peak information as the characteristic information of the RFI signal.

The first step involves selecting an RFI signal window. When extracting feature information from a signal segment containing RFI, it is necessary to select a time span that is as short as possible while containing at least one complete RFI period. The signal period is determined using the RFI signal period computation method proposed in Section 1.1. RFI signals have relatively large amplitude values compared to other signals, with clear start and end positions. The algorithm uses the position where the signal amplitude suddenly increases as the window starting point. The RFI signal feature extraction window size is related to the computed period. The interference signal window selection for a computer monitor is shown in Figure 2 [Figure 2: see original paper], with a window size of NMdMNNM181.

The second step involves taking the absolute value of the signal amplitude to obtain more peak information. The third step utilizes amplitude characteristics to acquire the peak information of the window. The fourth step preprocesses the obtained peak signal by replacing adjacent values with small amplitude differences with their mean value, thereby smoothing the peak curve, reducing local fluctuations, narrowing differences among signals of the same category, and improving recognition efficiency. Figure 3(a) shows the line chart of RFI signal peak information, which exhibits an unstable fluctuation trend that is not conducive to subsequent feature recognition. After smoothing, as shown in Figure 3(b), the curve becomes relatively stable and can reduce differences among similar signals. The smoothed peak information serves as the feature template for RFI signal recognition and enables similarity computation.

2 RFI Signal Feature Similarity Algorithm

After extracting features from an unknown signal, the similarity with known feature templates is calculated to achieve identification and classification of the unknown RFI signal. The algorithm employs a segmented scoring strategy to compute the similarity between two sequences. Each sequence can be divided into multiple segments based on adjacent discrete points. The algorithm compares whether the trends of corresponding segments in the two sequences are identical, awarding points for identical trends and deducting points for different trends.

Before similarity computation, it is necessary to ensure that the two sequences are in an "aligned" state, meaning that the segments of the two sequences correspond one-to-one. Since the specific details of the signal to be identified are unknown, the discrete peak information extracted is almost impossible to align with the feature template, meaning the phase information is not aligned. Consequently, the peak position information and other details need to be readjusted to calculate similarity.

Literature [13] proposed a method for measuring time series similarity called Dynamic Time Warping (DTW), which can compare time series of different lengths [14] and has been widely applied in speech signal processing. In this algorithm, a smaller Euclidean distance indicates higher similarity between two sound patterns [15]. The DTW algorithm can calculate the correspondence between discrete points of two sequences of different lengths when their Euclidean distance is minimized, as shown in Figure 4 [Figure 4: see original paper], which illustrates the correspondence between phase points of the two sequences.

Based on the DTW algorithm, this paper designs an RFI signal feature similarity computation algorithm with the following process: (1) Use the DTW algorithm to calculate the correspondence between discrete points of two sequence groups under the condition of minimum Euclidean distance. (2) Calculate the weight values of the discrete points of the unknown signal. (3) Compare segments: if the trends of corresponding segments are identical, the similarity score increases; if the trends differ, the score decreases. (4) Accumulate the scores of all segments, where the total score represents the similarity between the two sequences.

The first step of the RFI signal feature similarity computation algorithm uses the DTW algorithm to obtain the correspondence between discrete points of the sequences to be compared under minimum Euclidean distance, preparing for subsequent calculations. The second step calculates the weight values of the discrete points of the compared sequence using the method shown in Equation (1), where D represents the number of discrete points in the sequence, S(t) represents the amplitude value of the t-th discrete point, and W(t) is the weight value of the t-th discrete point. Larger discrete values have greater weights and stronger representativeness in signal features. To constrain the similarity value between -1 and 1, the weights of the first and last points are set to 0.

The third step computes sequence similarity using a segmented scoring accumulation method. The DTW algorithm calculates the correspondence of each segment, and segmented comparison is performed. As shown in Figure 4, the segments of the two sequences exhibit three situations: identical trends, different trends, and one-to-many correspondence. Points are added for identical trends as shown in Equation (2), where score represents the cumulative similarity score between the two sequences, M[i] is the weight of the i-th segment with a value equal to the average of the weights of the two discrete points in the segment (M[i] = 0.5 × (W[i] + W[i+1])). Points are deducted for different trends as shown in Equation (3). No points are added or deducted for one-to-many situations. After calculating and accumulating the scores of all segments, the similarity value between the two sequences is obtained.

Experimental Validation

The experiment selected interference sources provided on the ARRL website as experimental data, choosing seven RFI sources for cross-testing. The basic information is shown in Table 1 . To facilitate feature extraction and subsequent identification testing of RFI signals, the RFI signal period computation method proposed in Section 1.1 was used to obtain the period sampling points of each experimental dataset for the next steps of feature extraction and identification testing.

Ten periods of each signal were randomly selected as test data, and corresponding candidate feature templates were extracted for cross-testing with other signals. The candidate template with high similarity to the same signal and low similarity to other signals was selected as the optimal template. As shown in Figure 5 [Figure 5: see original paper], (a) selects k7_2 as the feature template for k7ing-3545khz, (b) selects ks_9 as the feature template for ks2am-streetlight, (c) selects mo_4 as the feature template for monitor, (d) selects n6_{10} as the feature template for n6rce-sps-carrier, (e) selects pl_8 as the feature template for plc-4, (f) selects au_7 as the feature template for ausoth, and (g) selects ot_4 as the feature template for 18120oth25.

The data from the group containing the optimal feature template for each signal were extracted, with results shown in Table 2 . The similarity of the same signal is the largest in its row, indicating that the extracted RFI feature templates can correctly identify the RFI source. The algorithm' s similarity computation is based on the feature template, and weight calculation depends only on the feature template. Therefore, when calculating the similarity between two sequences, the weight values are related only to the feature template, resulting in different values for symmetric positions across the diagonal. Differences within the same signal across different periods cause the similarity of the same signal to be less than 1.

Some values in Table 2 are relatively high, such as the value 0.726 in the n6rce-sps-carrier row and plc-4 column. Analysis reveals that this occurs because the amplitude fluctuation trends of the compared sequence pair are relatively consistent, leading to high similarity. Future work will continue to optimize the algorithm to improve recognition accuracy.

Conclusion

This paper designed a sliding convolution period computation algorithm that uses peak intervals after convolution to calculate the average signal period.

Based on the designed peak extraction algorithm, signal feature template extraction was implemented. A signal feature recognition algorithm was designed based on the DTW algorithm and scoring strategy, enabling identification and classification of unknown signals. Cross-correlation and autocorrelation calculations were performed on RFI data downloaded from the ARRL website. Experimental results demonstrate that the templates generated by the proposed algorithm exhibit significantly higher similarity with original signals than with other signals, indicating that the algorithm can effectively extract feature information and generate signal characteristic templates. The RFI feature recognition algorithm can correctly classify RFI. The proposed method enables fine-grained identification and labeling of RFI and is expected to provide a new solution for RFI feature recognition and labeling in radio astronomy.

The algorithm implementation code and experimental data for this paper have been open-sourced in a Gitee repository.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (11873082, 11803080, 12003062), the Youth Innovation Promotion Association of the Chinese Academy of Sciences, the National Key Research and Development Program (2018YFA0404704), the National Astronomical Data Center, and the Special Fund for Equipment Renewal and Major Instrument Operation of Astronomical Stations of the Chinese Academy of Sciences.

References

[1] R. NAN et al., "THE FIVE-HUNDRED-METER APERTURE SPHERICAL RADIO TELESCOPE (FAST) PROJECT," Int. J. Mod. Phys. D, vol. 20, no. 06, pp. 989–1024, 2011, doi: 10.1142/s0218271811019335.
[2] J.-M. Grießmeier, P. Zarka, M. Tagger, L. Collaboration, and others, "Radioastronomy with LOFAR," Comptes Rendus Phys., vol. 13, no. 1, pp. 23–27, 2012.
[3] E. J. Murphy, "A next-generation Very Large Array," Proc. Int. Astron. Union, vol. 13, no. S336, pp. 426–432, 2017, doi: 10.1017/s1743921317009838.
[4] J.-P. G. Porko, "Radio frequency interference in radio astronomy," 2011. Accessed: Nov. 30, 2020. [Online]. Available: https://aaltodoc.aalto.fi/handle/123456789/3739.
[5] Y. Wang et al., "Satellite {RFI} mitigation on {FAST}," Res. Astron. Astrophys., vol. 21, no. 1, p. 18, 2021, doi: 10.1088/1674-4527/21/1/18.
[6] J. Akeret, C. Chang, A. Lucchi, and A. Refregier, "Radio frequency interference mitigation using deep convolutional neural networks," Astron. Comput., vol. 18, pp. 35–39, 2017, doi: https://doi.org/10.1016/j.ascom.2017.01.002.
[7] W. A. Baan, P. A. Fridman, and R. P. Millenaar, "Radio Frequency Interference Mitigation at the Westerbork Synthesis Radio Telescope: Algorithms, Test Observations, and System Implementation," Astron. J., vol. 128, no. 2, pp. 933–949, Aug. 2004, doi: 10.1086/422350.

[8] A. R. Offringa, A. G. de Bruyn, M. Biehl, S. Zaroubi, G. Bernardi, and V. N. Pandey, "Post-correlation radio frequency interference classification methods," Mon. Not. R. Astron. Soc., vol. 405, no. 1, pp. 155–167, 2010, doi: 10.1111/j.1365-2966.2010.16471.x.
[9] U.-L. Pen et al., "The GMRT EoR experiment: limits on polarized sky brightness at 150 MHz," Mon. Not. R. Astron. Soc., vol. 399, no. 1, pp. 181–194, 2009, doi: 10.1111/j.1365-2966.2009.14980.x.
[10] D. Czech, A. Mishra, and M. Inggs, "A dictionary approach to identifying transient RFI," Radio Sci., vol. 53, no. 5, pp. 656–669, 2018, doi: 10.1029/2018RS006538.
[11] 吴鹏, 赵风海, 黄洋. 一种结合线性预测倒谱法和组合滑动窗口平滑法的基音周期估计改进算法 [J]. 南开大学学报 (自然科学版), 2019, 52(2): 29-33. Wu Peng, Zhao Feng-hai, Huang Yang. An Improved Pitch Period Estimation Cepstrum Algorithm Combined Linear Prediction and Sliding Window Smoothing[J]. Acta Scientiarum Naturalium Universitatis Nankaiensis(Natural Science Edition), 2019, 52(2): 29-33.
[12] 吴树兴, "一种语音信号基音周期时域估计算法," 电脑知识与技术, 2019,15(22):214-216. WU Shuxing.A time domain estimation algorithm for speech signal pitch period[J].Computer Knowledge and Technology,2019,15(22):214-216.
[13] T. K. Vintsyuk, "Speech discrimination by dynamic programming," Cybernetics, vol. 4, no. 1, pp. 52–57, 1968, doi: 10.1007/BF01074755.
[14] Müller, "Dynamic Time Warping," in Information Retrieval for Music and Motion, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 69–84.
[15] Y. Permanasari, E. H. Harahap, and E. P. Ali, "Speech recognition using Dynamic Time Warping ({DTW})," J. Phys. Conf. Ser., vol. 1366, p. 12091, 2019, doi: 10.1088/1742-6596/1366/1/012091.

Note: Figure translations are in progress. See original paper for figures.