Ensemble Feature Selection and Evaluation for Single-Pulse Signal Classification Postprint

Authors: Zhang Jinqu, Ling Yu, Du Ping, Li Xiangru, Li Hui

Date: 2023-10-07T00:00:00+00:00

Abstract

Affected by numerous radio frequency interference signals, the rapid and accurate identification of single-pulse signals from massive observational data has become a critical task in astronomical data processing. Designing and extracting effective data features is the decisive factor for efficiently identifying single-pulse signals using machine learning. To address the key challenge of selecting optimal features to improve classification accuracy for single-pulse signals, we propose an ensemble feature selection method tailored for single-pulse signal classification. The method first combines parametric, statistical, and abstract features of single-pulse signals, then employs five individual feature selection methods to obtain their respective optimal feature subsets, and finally applies a greedy strategy to ensemble-screen these subsets to derive the optimal integrated feature set. Experimental results demonstrate that the optimal feature set comprises both statistical and abstract features. Under the same feature dimensionality, ensemble feature selection achieves higher model accuracy compared to individual methods, yielding up to a 1.8% improvement in F1 score. In the context of massive data, ensemble feature selection plays a crucial role in reducing feature dimensionality, enhancing classification performance, and accelerating data processing speed.

Full Text

Ensemble Feature Selection and Evaluation for Single Pulse Signal Classification

Jinqu Zhang[1]†, Yu Ling[1], Ping Du[2], Xiangru Li[1]‡, Hui Li[1]

[1]School of Computer Science, South China Normal University, Guangzhou 510631
[2]School of Building Information, Guangdong Construction Vocational Technology Institute, Qingyuan 511500

Abstract

Affected by a large number of radio frequency interference signals, the rapid and accurate identification of single pulse signals from massive observational data has become a critical task in astronomical data processing. Designing and extracting effective data features is the decisive factor for efficient identification of single pulse signals using machine learning. To address the key problem of how to select optimal features to improve classification accuracy for single pulse signals, this paper proposes an ensemble feature selection method. The method first combines parametric features, statistical features, and abstract features of single pulse signals, then uses five individual feature selection methods to select their respective optimal feature subsets, and finally employs a greedy strategy to integrate and filter the optimal subsets obtained from the five methods to acquire the best ensemble feature set. Experiments demonstrate that the optimal feature set contains both statistical and abstract features. With the same number of features, ensemble feature selection achieves higher model accuracy than individual methods, improving the F1-score by up to 1.8%. In the context of massive data, ensemble feature selection plays an important role in reducing feature dimensionality, improving classification performance, and accelerating data processing speed.

Keywords: pulse signal, radio pulsar, action variable, methods: data analysis

1 Introduction

Single pulse signals refer to pulsed radiation signals emitted by celestial bodies without fixed periods, primarily divided into two categories: Rotating Radio Transients (RRATs) and Fast Radio Bursts (FRBs) [1–3]. With continuous technological development and improving sensitivity of astronomical observation equipment, the received pulse signals contain increasingly more interference signals. Radio frequency interference signals affected by aircraft, radar, ionosphere, and other sources are growing exponentially. How to accurately identify celestial single pulse signals from massive observational data has become an important task in astronomical data processing. To this end, scholars have conducted extensive research. Currently, machine learning-based methods have become the primary approach for single pulse signal mining, and how to design and extract pulse signal features is a key factor affecting machine learning performance [4]. By screening effective features, redundant features can be removed, reducing computational load for data processing to some extent while improving recognition accuracy. This is particularly helpful for enhancing single pulse signal search efficiency in the context of high-speed, large-scale sky surveys.

Based on their sources and calculation methods, pulse signal features are mainly divided into three categories: parametric features, statistical features, and abstract features. Parametric features are determined by the signal receiver, spatial environment, and data processing pipeline when receiving pulse signals. For example, Dispersion Measure (DM) is the integrated column density of free electrons between a celestial body and Earth along the signal propagation direction, with units of $pc \cdot cm^{-3}$. Although determined by the spatial environment, DM has significant impact on pulse signal classification and recognition, making it a typical parametric feature. Similarly, Signal-to-Noise Ratio (S/N) is the ratio of the voltage value of signals received by radio telescopes to the noise voltage recorded simultaneously. Higher S/N indicates relatively stronger signal intensity compared to noise, and S/N is also a primary basis for identifying pulse signals. Parametric features are typically recorded directly in data documents during initial reception and processing of celestial signals and can be obtained through simple calculations later. Their characteristics include simple acquisition, clear meaning, and good classification effectiveness for pulse signals.

Statistical features refer to quantitative features with descriptive significance that are manually designed through observation and calculation of data. For instance, Lyon et al. calculated four unbiased statistical features based on pulse profile curves and DM-S/N curves: mean, standard deviation, excess kurtosis, and skewness, which demonstrate good performance in single pulse signal classification [5]. Building upon Lyon' s unbiased statistical features [5], Tan et al. added relevant statistical features based on time-phase diagrams, frequency-phase diagrams, and pulse profile diagrams, significantly reducing the false positive rate in classification [6]. Statistical features have clear meanings but their design is heavily influenced by experience and may omit important statistical features.

Abstract features refer to features automatically extracted by algorithms without manual design. Currently, convolution operations based on Convolutional Neural Networks (CNN) are the most commonly used abstract feature extraction methods. Using different convolution kernels through multiple layers of convolution operations ultimately outputs a series of features that have no explicit meaning but demonstrate good classification effects. This approach is becoming mainstream across various fields and plays an increasingly important role in single pulse signal recognition. For example, Zhu et al. designed a pulsar classification system based on images called PICS (Pulsar Image-based Classification System). This system screens pulsar signals through four sub-images output by PRESTO (PulsaR Exploration and Search Toolkit) software, uses CNN to automatically learn pulsar features from pulsar candidates, and then employs classification algorithms such as Support Vector Machine, Artificial Neural Networks (ANN), and logistic regression for pulsar signal classification [7]. Wang et al. proposed the PICS-ResNet model based on the PICS system, primarily by replacing the original CNN with ResNet (Residual Networks). Experiments on observational data from FAST (the Five-hundred-meter Aperture Spherical radio Telescope) and GBNCC (Green Bank North Celestial Cap) achieved higher classification performance [8]. In 2020, Agarwal et al. built 11 deep learning models based on features extracted from eight deep network architectures such as VGG (Visual Geometry Group) and Densenet (Dense Convolutional Network), detecting over 2000 single pulse signals from more than 20 pulsars [9-

10].

Applications show that abstract features based on CNN can effectively classify and recognize pulse signals, but they suffer from poor interpretability and unclear meanings. Additionally, features extracted using CNN often contain redundant features, which not only consume computational resources but also affect classification accuracy to some extent. Therefore, how to fully utilize the respective advantages of parametric, statistical, and abstract features is of great significance for single pulse signal classification. The goal of this paper is to design an integrated feature selection and evaluation method that provides feature selection solutions for machine learning-based single pulse signal classification.

2 Data Sources

This study directly uses the annotated single pulse dataset from Michilli et al. [11] for experimental analysis. This dataset originates from the LOFAR tied-array all-sky survey (LOTAAS) project, with specific formation details available in references [11-12]. The dataset contains 3.74 million pulse signal records belonging to 53,066 pulse events, including 35,063 radio frequency interference events and 18,003 pulse events from 47 known pulsars. Signal events belonging to the same pulse form a dispersed pulse group.

3 Ensemble Feature Selection Method Design

The basic idea of ensemble feature selection is to select the optimal feature combination most suitable for single pulse signal classification from a feature set composed of parametric, statistical, and abstract features. The overall technical process is shown in Figure 1 [Figure 1: see original paper], mainly divided into three steps: (1) calculating parametric, statistical, and abstract features respectively to form a multi-source original feature set; (2) using individual feature selection methods to extract optimal feature subsets from the mixed feature set; and (3) using a greedy strategy to screen the optimal ensemble feature subset from multiple optimal feature subsets.

3.1 Feature Design

3.1.1 Parametric and Statistical Feature Design Based on the characteristics of pulse signal data and combining feature designs from existing studies [11-12], the parametric and statistical features used in this paper are shown in Table 1 .

3.1.2 Abstract Feature Extraction Based on Convolutional Neural Networks Convolutional Neural Networks can perform multi-layer feature extraction from different receptive fields through convolution, activation, pooling, and other operations, achieving successful applications in image classification and recognition. Using CNN to extract abstract features from data distribution diagrams of single pulse signals will greatly enhance the feature sources for single pulse signals. This paper builds a deep residual shrinkage network and uses the signal-to-noise ratio and window width distribution curve morphology images of each dispersed pulse group as network model inputs to extract abstract features of single pulse signals.

The deep residual shrinkage network designed in this paper, called RSDFNet (Residual Shrinkage Distribution curve Feature extraction Network), is based on the deep residual neural network proposed by He et al. [13] and introduces residual modules on top of CNN. The model structure is shown in Figure 2 [Figure 2: see original paper]. The last hidden layer of RSDFNet serves as the feature extraction layer to obtain abstract features learned from the distribution curve morphology images of signal-to-noise ratio and window width.

3.2 Individual Method-Based Feature Subset Construction

The multi-source mixed feature set inevitably contains numerous redundant and invalid features, which not only reduce model computational efficiency and cause dimensionality disasters but also affect model accuracy. Therefore, screening out the most useful features is significant for model computation. However, there are many different methods for evaluating the importance of a feature for classification tasks. This paper first uses individual feature selection methods including Chi-square test [14], Mutual Information [15], Recursive Feature Elimination [16], and embedded feature selection to screen out the optimal feature subset for each method. Then, the optimal feature subsets from multiple methods are integrated and filtered to form the optimal ensemble feature combination, achieving complementary advantages of various feature selection methods.

3.2.1 Chi-square Test-Based Feature Subset Selection The basic idea of the Chi-square test is to determine the correctness of theoretical values by observing deviations between actual and theoretical values. The specific approach is to first assume that two variables are independent (the "null hypothesis"), then observe the deviation degree between actual values (observed values) and theoretical values. If the deviation is sufficiently small, the two are considered truly independent, and the null hypothesis is accepted. If the deviation is large enough, the two are considered related, rejecting the null hypothesis and accepting the alternative hypothesis. For single pulse signal feature selection, the null hypothesis "the extracted feature is not correlated with the single pulse signal to be identified" is used. A larger calculated Chi-square value indicates greater deviation from the null hypothesis, suggesting that the opposite of the null hypothesis is correct. Therefore, a larger Chi-square value indicates higher correlation between the feature and single pulse signal. The Chi-square calculation formula is shown in equation (1):

$$\chi^2 = \sum \frac{(A-E)^2}{E}$$

where A is the actual value calculated based on a feature, and E is the theoretical value.

3.2.2 Mutual Information-Based Feature Subset Selection Mutual Information can measure the mutual dependence between two random feature variables [15] and is typically used to evaluate the information contributed by the occurrence of one event to another. In classification, it can be viewed as the contribution of a certain feature to distinguishing a certain class. When variables X and Y are discrete random feature variables, the calculation formula is:

$$I(X,Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}$$

In the formula, $I(X,Y)$ represents the mutual information between X and Y, $p(x,y)$ is the joint probability distribution function of X and Y, and $p(x)$ and $p(y)$ are the marginal probability distribution functions of X and Y respectively. If the mutual information value is zero, it indicates that the two random variables provide no information to each other and are independent. A larger mutual information value indicates higher dependence between the two variables.

3.2.3 Recursive Feature Elimination-Based Feature Subset Selection Recursive Feature Elimination trains a model on a given feature set, removes the least important features based on model results, and continues training on the remaining feature set. This process repeats until the number of features in the set reaches a specified value, thereby selecting the optimal feature subset [16]. In this paper, the LightGBM (Light Gradient Boosting Machine) model is selected for recursive feature elimination. LightGBM is a gradient boosting framework based on decision trees, optimized from the traditional GBDT (Gradient Boosting Decision Tree) algorithm. It supports multi-threaded parallel computing, reduces memory consumption while ensuring accuracy, and greatly improves training speed, thereby achieving efficient processing of massive data [16].

3.2.4 Embedded Feature Selection Embedded feature selection integrates feature data with the model for a given base learner, filtering out features with zero coefficients during model training. It has low computational cost, fast feature selection speed, and can greatly reduce data dimensionality. This paper selects Random Forest and XGBoost (Extreme Gradient Boosting) learners as base models for feature selection [17-18]. Both embedded learners can effectively model nonlinear relationships between features. During feature selection, the models calculate feature correlation coefficients and contribution metrics to model performance. When the correlation coefficient or contribution metric falls below a set threshold, the feature is automatically discarded.

3.3 Greedy Strategy-Based Ensemble Feature Selection

Individual feature selection methods cannot comprehensively evaluate data features, and comprehensively utilizing the advantages of multiple feature selection methods is an effective way to compensate for the limitations of individual methods. Therefore, this paper proposes a greedy strategy-based ensemble feature selection method with the following specific steps:

(1) Features extracted by each individual method are sorted by importance from large to small. Assume the feature subset given by the i-th method is $S_i = \{s_{i,j} | i = 1, 2, 3, 4, 5; j = 1, 2, 3, \cdots, m\}$, where $s_{i,j}$ represents the feature ranked j in the feature subset given by the i-th method, m represents the total number of features in the subset, and the total number of methods is n.

(2) Take the top-ranked feature from each feature subset and place them into a buffer collection B. After removing duplicates from B, input each feature individually into the LightGBM classification model to obtain corresponding classification performance. Select the feature with the best classification performance, denoted as c_1. Remove c_1 from B and add it to the ensemble feature set C.

(3) Take the second-ranked features from each feature subset, i.e., $s_{i2}, i = 1, 2, 3, 4, 5$. Place these newly selected five features into collection B and remove duplicates. Then, combine each element from B with the previously selected optimal feature c_1 and input them into the LightGBM classification model to obtain corresponding classification performance. Select the feature combination with the best performance, denote the second selected feature as c_2, remove it from B, and add it to the ensemble feature set C.

This process continues iteratively to select features $c_3, c_4, \cdots, c_m$ until all features in the feature subsets are screened. Finally, an ensemble feature set $C = \{c_i | i = 1, 2, 3, \cdots, m\}$ sorted by feature importance is obtained. The algorithm flow of the ensemble feature selection method is as follows:

Algorithm: Ensemble Feature Selection Method

Input: Ordered feature set list S_i; Number of single feature selectors n; Number of features m
Output: Selected feature set C using ensemble feature selection method

1. Initialize temporary collection B and results feature collection C

2. for $i = 1$ to n:
   3. for $j = 1$ to m:
   4. $B \leftarrow [B; s_{ij}]$
   5. end for
   6. Remove duplicate features in B
   7. for $k = 1$ to size(B):
   8. Get out the k-th feature b_k
   9. if b_k is not in C:
   10. Compute classification performance of subsets $\{c_1, \ldots, c_{i-1}, b_k\}$
   11. end if
   12. end for
   13. Record the best performance feature b_l based on combination of b_k and C
   14. $C \leftarrow [C; b_l]$

3. end for

4. return C

4 Results and Discussion

During experiments, dispersed pulse groups belonging to single pulses and radio frequency interference were randomly divided into training, validation, and test sets at a ratio of 6:2:2. The F1-score, which combines precision and recall, was used as the primary evaluation metric. The main experimental results are as follows.

4.1 Classification Performance Analysis of Different Neural Network Models

With the development of Convolutional Neural Networks, an increasing number of network models have emerged. This paper selected several representative network models for single pulse signal classification performance comparison to determine the optimal network structure for abstract feature extraction. During experiments, each network was individually optimized through automatic parameter search to obtain the best performance. The experimental results for each model are shown in Table 2 .

The results show that the RSDFNet model used in this paper achieves an F1-score of 95.3%, demonstrating the best overall performance among these models and proving RSDFNet' s strong capability to learn and extract features from the distribution curve morphology of signal-to-noise ratio and window width. Compared with ResDFNet, RSDFNet' s F1-score improved by 0.4% after introducing the residual shrinkage module. Analysis suggests that since signal events in dispersed pulse groups are grouped based solely on the proximity of signal time and dispersion values from each record in the signal event table without considering their correlations, the distribution curve morphology diagrams tend to contain features of non-correlated noise event points. RSDFNet uses an attention mechanism to focus on these unreasonable feature points in the distribution curve morphology images and soft thresholding to set them to zero, thereby strengthening the model' s feature extraction capability from these distribution curve morphology diagrams.

4.2 Impact of Feature Quantity Extracted by Convolutional Neural Networks on Single Pulse Signal Classification Results

This paper primarily extracts abstract features contained in input images through the last hidden layer of the RSDFNet model. The number of nodes in this layer also affects model classification performance. Therefore, by adjusting the number of abstract features (i.e., adjusting the number of nodes in RSDFNet' s last hidden layer), we observed model classification performance to seek higher-quality features. The experimental results are shown in Figure 3 [Figure 3: see original paper].

The figure intuitively shows that as the number of abstract features increases, model performance improves accordingly. When the number of abstract features reaches 16, the model' s F1-score is highest. Thereafter, as the number of features continues to increase, model performance no longer improves but instead declines continuously. Therefore, this paper sets the number of extracted abstract features to 16.

4.3 Comparison Between Individual and Ensemble Feature Selection Results

Based on the individual and ensemble feature selection methods described in Section 3, all mixed features of single pulse signals were screened. The feature importance rankings obtained by different feature selection methods are shown in Table 3 .

In Table 3, features named with "f+number" are abstract features extracted based on the deep residual shrinkage network, while features named otherwise are parametric and statistical features. The table shows that the feature importance rankings calculated by each method differ significantly. The Mutual Information method selects statistical and parametric features as important features, while the Random Forest-based embedded feature selection method treats abstract features as important. From the ensemble feature selection results, abstract feature f9 is the most important feature, followed by statistical and parametric features. Overall, relying solely on one type of feature—for example, using only statistical features or only abstract features from the deep residual shrinkage network—does not yield the best feature set. Integrative selection from multi-source features is an effective method for constructing the optimal feature set.

4.4 Analysis of Feature Quantity Impact on Model Performance

Although the previous section obtained feature importance rankings under different methods, the optimal number of input features for achieving the best classification results remains uncertain. Therefore, this section further discusses the impact of input feature quantity on model performance. Using the LightGBM model as an example and the F1-score as the evaluation metric, we calculated the F1-score of the model under different numbers of input features. LightGBM was chosen because it offers faster training speed and higher efficiency compared to other models like XGBoost, and is suitable for large-scale data processing [19]. Additionally, LightGBM is essentially a tree-based model with numerous hyperparameters that affect tree structure, training speed, and model fit. These hyperparameters also interact with each other—for example, the parameter num_{leaves} affects both decision tree structure and fitting degree, while max_{bin} is related to efficiency, accuracy, and fitting degree. Therefore, manual parameter adjustment should be avoided, and automatic search methods are preferred for determining hyperparameters. This paper selected eight commonly used hyperparameters for the LightGBM model and used the Sparrow Search Algorithm for automatic adjustment. These eight parameters and their search ranges are shown in Table 4 .

To analyze the impact of feature quantity on model performance and compare the performance of individual and ensemble feature selection methods, comparative experiments were conducted with the same number of features. According to the feature importance rankings in Table 3, feature subsets for different feature selection methods were constructed from smallest to largest importance. These subsets were input into the LightGBM model for training and classification prediction. The F1-scores of five individual feature selection methods were calculated based on classification results, and the maximum F1-score among individual methods and the F1-score of the ensemble feature subset were compared for each feature quantity. The variation with feature quantity is shown in Figure 4 [Figure 4: see original paper].

The figure shows that as the number of input features increases, the F1-score rises rapidly, reaching its maximum for the ensemble method at approximately 8 features and for individual methods at approximately 10 features. Subsequently, as the number of features continues to increase, the F1-score plateaus and slightly decreases. This indicates that additional features may be redundant or invalid, demonstrating that more features are not necessarily better for single pulse classification.

The comparison between individual and ensemble feature selection shows that under the same number of features, the ensemble feature selection method consistently outperforms individual methods. The ensemble feature subset achieved a maximum F1-score of 99.2%, improving upon individual methods by up to 1.8% at the same feature quantity, demonstrating the effectiveness of the ensemble feature selection method. The ensemble approach combines results from multiple individual methods, making it easier to identify features with strong discriminative capability.

4.5 Performance Gain Analysis of Abstract Features on Different Models

Based on previous analyses, the optimal feature set selected by the ensemble method contains three abstract features and five custom features. This indicates that while some empirically designed statistical features are effective, relying solely on manual features may not achieve optimal results. This section further analyzes the performance gain of abstract features extracted by neural networks on different models. We compared the classification performance of single pulse signals on SVM (Support Vector Machines), KNN (K-Nearest Neighbors), AdaBoost (Adaptive Boosting), and LightGBM models using only the manual features from Table 1 versus combining them with 16 abstract features extracted by RSDFNet. Model parameters were optimized through automatic search for each model. The experimental results are shown in Table 5 .

Table 5 shows that after adding abstract features, most models exhibit corresponding improvements in accuracy and F1-score, with the greatest improvement observed for the KNN model (F1-score increased by up to 15%). Although KNN shows the largest improvement in accuracy and F1-score, LightGBM maintains the highest accuracy and F1-score both before and after adding abstract features. The SVM model' s accuracy did not improve but slightly decreased, possibly because the SVM classification boundary can be constructed with few features, or because abstract features contain some invalid or redundant features. Through ensemble feature selection, the optimal feature combination can be further screened.

5 Conclusion

Machine learning has become the primary method for single pulse signal detection and identification, and feature extraction for pulse signals has become an important aspect affecting machine learning effectiveness. To this end, this paper designs an ensemble feature selection method based on parametric features, statistical features, and abstract features. The method first uses Chi-square test, Mutual Information, Recursive Feature Elimination, and embedded feature selection methods to screen optimal feature subsets from different perspectives, then employs a greedy strategy to select the final feature combination for classification from the optimal feature collection.

Based on experimental result analysis, different feature selection methods produce significantly different feature importance rankings, and feature selection methods have a clear impact on classification accuracy. When the number of features is small, different feature selection methods have a greater impact on classification results. When the number of features exceeds 10, classification performance of different feature screening methods begins to converge. Compared with individual feature selection methods, the F1-score based on ensemble features can be improved by 1.8%, indicating that ensemble feature selection provides better improvement for single pulse classification accuracy.

From the composition of ensemble features, the features selected by the ensemble method include abstract features extracted by neural networks, parametric features, and statistical features. This demonstrates that relying solely on abstract features from CNN or solely on manually designed statistical features is difficult to achieve optimal classification results. Mixed application of multi-source features is an effective means to improve single pulse signal classification. This work provides a new perspective for machine learning-based single pulse signal classification. By selecting effective ensemble features, not only is the number of features reduced, but classification performance is also improved. The reduction in feature number further decreases computational load for model data processing, which is of great significance for improving the efficiency of massive astronomical data processing in the context of high-speed, large-scale sky surveys.

References

[1] Cordes J M, McLaughlin M A. ApJ, 2003, 596: 1142
[2] McLaughlin M A, Lyne A G, Lorimer D R, et al. Nature, 2006, 439: 817
[3] Lorimer D R, Bailes M, McLaughlin M A, et al. Science, 2007, 318: 777
[4] Wang Y C, Zheng J H, Pan Z C, et al. Journal of Deep Space Exploration, 2018, 5: 203
[5] Lyon R J, Stappers B W, Cooper S, et al. MNRAS, 2016, 459: 1104
[6] Tan C M, Lyon R J, Stappers B W, et al. MNRAS, 2018, 474: 4571
[7] Zhu W W, Berndsen A, Madsen E C, et al. ApJ, 2014, 781: 117
[8] Wang H F, Zhu W W, Guo P, et al. SCPMA, 2019, 62:
[9] Agarwal D, Aggarwal K, Burke-Spolaor S, et al. MNRAS, 2020, 497: 1661
[10] Agarwal D, Lorimer D R, Surnis M P, et al. MNRAS, 2020, 497: 352
[11] Michilli D, Hessels J W T, Lyon R J, et al. MNRAS, 2018, 480: 3457
[12] Ling Y, Zhang J Q, Li X R, Li H. Astronomical Research & Technology, 2022, 19: 264
[13] He K, Zhang X, Ren S, et al. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016. Las Vegas, 2016: 770
[14] Kumar V, Minz S. Smart Computing Review, 2014, 4:
[15] Hira Z M, Gillies D F. Adv Bioinformatics, 2015, 2015:
[16] Albawi S, Abed Mohammed T A, Al-Zawi S. 2017 International Conference on Engineering and Technology (ICET), August 21-23, 2017. Antalya, 2018: 1
[17] Kwak N, Choi C H. ITPAM, 2002, 24: 1667
[18] Chen T, Guestrin C. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: Association for Computing Machinery, 2016: 785
[19] Ke G, Meng Q, Finley T, et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree//Guyon I, Von Luxburg U, Bengio S, et al. Advances

in Neural Information Processing Systems. NewYork: Curran Associates Inc., 2017: 3146

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