PAPER • OPEN ACCESS

Super Base Station Fault Detection Mechanism Based on Negative Selection Algorithm and Expert Knowledge Base

To cite this article: Guanwen Ye et al 2019 IOP Conf. Ser.: Mater. Sci. Eng. 490 072019

View the article online for updates and enhancements.
Super Base Station Fault Detection Mechanism Based on Negative Selection Algorithm and Expert Knowledge Base

Guanwen Ye\textsuperscript{1,2,*}, Yuanyuan Wang\textsuperscript{2}, Qian Sun\textsuperscript{2}

\textsuperscript{1}Chongqing University of Posts and Telecommunications, Chongqing, China; \textsuperscript{2}Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

*Corresponding author e-mail: 1024096567@qq.com

Abstract: With the expansion of network scale and the increase of equipment complexity, the super base station business data has experienced explosive growth, which makes fault detection more and more difficult, and the efficiency of fault management is getting lower and lower. To solve the above problems, this paper designs a comprehensive fault detection mechanism (NSAEFD) by combining the negative selection algorithm in the field of artificial immunity and expert system. NSAEFD is further introduced in detail in super base station (SBS). NSAEFD is also easy to implement and can be well applied to the fault management system of the SBS.

1. Introduction

With the rapid development of mobile Internet, online voice, video, real-time mobile games, mobile phone shopping and other services, business data traffic index of the base station has increased. Traditional cellular mobile communication systems cannot meet the requirements of service and energy consumption in the future. Therefore, the Institute of Computing Technology of the Chinese Academy of Sciences has proposed the concept of a super base station (SBS)\textsuperscript{[1-3]}. SBS is an access network platform of multi-mode heterogeneous with physical concentration and logical distribution, which supports horizontal sharing and statistical multiplexing of resources. SBS create a large number of virtual base stations (VBS) in a centralized large-scale resource pool through a unified and open interface, and multiple VBS share resources in the resource pool. Once a problem occurs in the resource pool, it may cause lots of VBS failures associated with it, thereby affecting a large scale of access user services, even causing the entire network to crash. Therefore, SBS requires a more secure and reliable network environment and a more efficient fault management system.

Fault management includes fault detection, fault location, and fault repair. As the first step of fault management, fault detection is the basis of fault location and fault repair. The performance of fault detection directly affects the effect of fault management and is critical to fault management. To cope with the future large-capacity, high-bandwidth, and multi-type service scenarios, the fault detection of SBS needs to possess the ability to detect potential abnormalities in advance, reduce the probability of failure and improve the reliability. This paper proposes a fault detection mechanism (NSAEFD) based on the combination of the negative selection algorithm and expert knowledge base. NSAEFD divides into two steps: 1. Abnormal detection: firstly, use the negative selection algorithm to generate the anomaly detectors. Next, find the abnormal condition by anomaly detections. 2. Fault determination: filter the abnormal by the expert system rule base. If the abnormality is known and causes a fault, solve the problem by expert knowledge base; If the abnormality is unknown, notify the fault management module to perform fault location, and add the solutions to the expert system.
2. Technology overview

2.1. Anomaly Detection Methods

At present, the main anomaly detection methods are using the data mining, neural networks, artificial immune systems, etc. [4] proposed using data mining technology to train fault data and generate a set of rules for fault detection. [5] proposed a fast convergence algorithm based on fuzzy inference to detect the law of network signals, and use this law to judge whether the network transmission signal error caused network failure. The above methods could effectively solve the massive data scenarios, but need to train the fault samples. The larger the fault samples, the higher the training rules and model reliability, but for most devices, especially in SBS, it is difficult to get a large number of fault samples at a time. Therefore, the above anomaly detection methods are not suitable for SBS. [6] proposed a negative selection algorithm based on the artificial immune system, which draws on the "negative selection" process when immune cells mature and learns the abnormalities learned by learning "self" data. The detector is also used to detect the "non-self" situation, wherein the "self" data refers to normal data, and the "non-self" refers to abnormal conditions. This method is simple to implement and only needs to train normal data to detect anomalies. The negative selection algorithm includes two phases, the offline training phase, and the online detection phase.

2.1.1. Offline Training Phase

The purpose of the offline training phase is to generate an anomaly detector for anomaly detection. It mainly includes the following steps:

(1) Collecting self-samples;
(2) Randomly generating a candidate detector;
(3) Matching the self-sample with the candidate detector. If the matching is satisfied, the candidate detector is deleted and returned (2); Otherwise, the candidate detector becomes a new one. An anomaly detector added to the mature detector set;
(4) The training ends when enough anomaly detectors are reaching.

2.1.2. Online Detection Phase

Online detection phase is to filter out abnormal data from the untested data. It mainly includes the following steps:

(1) Match the test sample to the mature anomaly detector;
(2) If it is satisfied, an abnormal sample is detected; if not match, return (1);
(3) Until all detectors in the anomaly detector set participate in the detection.

2.2. Expert System

The expert system[7] is mainly composed of components such as knowledge base, inference engine, and human-computer interaction interface.

The knowledge base stores a great deal of expertise and experience, and the form of knowledge expression are generally as follows: IF <condition> or <premise>, THEN < conclusion> or <operation>, SOLUTION <method>. Use connectors such as AND and OR in the middle. The connector is the rule element, which is a description of the alarm information and solutions.

The inference engine is the organization control mechanism of the expert system. It uses the knowledge in the knowledge base according to the input information and performs reasoning according to a certain strategy to complete the fault diagnosis.

The human-machine interface is the medium for information exchange between the expert system and the user. Through the customized interface, engineers can easily add, delete and modify expert knowledge. This effectively extends and maintains a knowledge base of expert systems.
3. Design of NSAEFD

3.1 Main Idea

The design of NSAEFD in SBS divides into two steps: 1. Abnormal detection; 2. Fault determination. As is shown in Figure 1. Firstly, the normal operation data of the base station is trained to generate anomaly detectors, and then it is responsible for supervising the data generated when the ABS is running and performing abnormality detection. If an abnormality is detected, then start model matching by the expert system. If it is a known abnormality, repair according to the expert knowledge base; If an unknown abnormality is detected, notify the fault management system for fault location, and add the fault case to the expert system.

![Figure 1. Fault Detection Overview](image)

3.2 Build Expert Knowledge Base

The expert knowledge base built before fault detection. The sources of expert knowledge are as follows: (1) Common standards and indicators proposed by communication equipment manufacturers or mobile communication protocol developers. (2) SBS operation and maintenance experts combined with the previous fault cases of products to summarize the knowledge. (3) The potential knowledge mined by the fault detection mechanism, that is, the new knowledge collected before this test.

The abnormal information and solution of the SBS stored in the knowledge base according to the expression form of the expert system. For example, the cell handover fails, or the call drop rate is high, because the handover parameter setting is unreasonable, and the solution has an adjustment handover threshold, delay, antenna tilt angle, and so on.

\[
\text{IF 'cell_handoff' = fail OR 'drop_call_rate' \geq 0.10}
\]

\[
\text{THEN 'handoff_parameter' = unreasonable SOLUTIONS 'measure A or measure B or...'}
\]

The inference engine designs in 2 steps:

1. Model matching: Match the currently detected abnormal conditions with the knowledge base. If it matches exactly or roughly matches, step (2) is triggered.

2. Competitive solution: Select the most qualified method from the solution strategy and provide it to the management terminal.

3.3 Abnormal Detection

3.3.1 Data Pre-processing

The performance parameters of the SBS include: delay, packet loss rate, handover success rate, dropped call rate, cell interference threshold, etc. The parameters are not a dimension level.[8] proposed to normalize different kinds of data into real-value vector sequences, normalized formulas (3-1)given, the anomaly detector is also represented by a sequence of the real-value vectors. The distance between the two real-value vectors indicates the affinity. For example, the vector \(x=[0.3, 0.5, 0.1, 0.4]\) represents the untested sample, and the vector \(y=[0.4, 0.6, 0.2, 0.3]\) represents the detector. when the distance between the two vectors is smaller than a certain threshold, it means that the sample matches the detector. The affinity between the vector x and the vector y in this example is calculated using the Euclidean distance formula (3-2).

\[
x_{\text{dist}} = \frac{x - y}{||x|| \cdot ||y||}
\]

where \(x\) and \(y\) represent the bits of the vectors \(x\) and \(y\), respectively. \(n\) represents the dimension
of the real value vector, which is the number of parameter types.

\[ d_o = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \]  (3-2)

Where \( x \) is the raw data, \( x_{\text{max}} \) and \( x_{\text{min}} \) are the maximum and minimum values of the original data set, and \( x_{\text{norm}} \) is the normalized data, corresponding to each element in the real value vector.

### 3.3.2 Generating Anomaly Detector

Take enough normal operating parameters from SBS log and normalize them to a set of real-valued vectors through Section 3.3.1. We call them self-samples. The affinity radius of the self-sample is \( r_s \), which indicates the affinity range of the self-sample. If the minimum distance between the untested point and the self-sample is greater than the absolute distance, the point to be tested belongs to the abnormal data; otherwise, the point to be tested belongs to the normal data. The generation process of the anomaly detector is a negative selection process: randomly generate a detector sequence, compare the minimum distance \( d_{\text{min}} \) between the detector and the self-sample, and \( d_{\text{min}} \) is calculated by the formula (3-2). If \( d_{\text{min}} < r_s \), the detector is negated. If \( d_{\text{min}} > r_s \), the detector can be used as a candidate detector, and the corresponding detection radius is \( L = d_{\text{min}} - r_s \). The detection radius set of the existing anomaly detectors \( r_i \), and \( i \) is the number of anomaly detectors. To reduce the coincidence rate between the detectors, it is necessary to judge the size of \( L \) with \( r_i \). If \( L < r_i \) exists, then the candidate detector will be discarded; Otherwise, the candidate detector will be added to the mature detector set; Figure 2-a is a flow chart of the anomaly detector generation phase.

The stopping criterion for generating a detector set is the value at which the detector reaches a predetermined coverage. [9] proposed to use the sample to estimate whether the coverage has been reached, and to temporarily stop generating the detector when performing sample estimation. Select \( n \) test samples, let \( x \) is the number of test samples covered by the detector, and \( \hat{p} \) in formula (3-3) is the estimated coverage. \( p \) is the predetermined coverage rate, and \( \sigma \) is the standard deviation. According to the central limit theorem, when the test sample \( n \) is large enough, the error \( z \) value (3-4) of the coverage of the test sample estimate can be approximated to obey the standard normal distribution.

\[ \hat{p} = \frac{x}{n} \]  (3-3)

\[ z = \frac{\hat{p} - p}{\sigma / \sqrt{n}} \]  (3-4)

\[ z = \frac{x}{\sqrt{npq}} \]  (3-5)

(3-3) and (3-4) can be derived (3-5). However, the estimate would produce the error. [10] proposed when \( z > z_\alpha \), it can be considered that the coverage has been reached, and stop the training; when \( z < z_\alpha \), the predetermined coverage range is not reached, and continue generating the detectors. \( \alpha \) is a significant level, and the smaller the \( \alpha \), the more accurate the result of reaching the predetermined coverage. Figure 2-b is the process for determining the predetermined coverage.

The abnormal detector generation steps of SBS are as follows:

1. Select the normalized real-valued sample in Section 3.3.1 and set the self-samples radius \( r_s \).
2. Randomly generate a detector to calculate the minimum distance \( d_{\text{min}} \) to self-samples.
3. Compare \( d_{\text{min}} \) and \( r_s \). If \( d_{\text{min}} \) is less than \( r_s \), the random detector belongs to the self-samples and returns to step (2). Otherwise, the detector is listed as a candidate detector and then calculate the candidate detector radius \( L = d_{\text{min}} - r_s \).
4. Comparing \( r \) with \( r_i \) of all existing detectors, if \( r \) is less than \( r_i \), it coincides with detector i, discarding the candidate detector; if \( r \) is greater than all current \( r_i \), adding the candidate detector to the the mature detection Collection.
5. Next, judge whether the mature detector is sufficient (enter in Figure 2-b), and stop generating the detector at this time.
6. Select the significance level \( \alpha \), the scheduled coverage rate \( p \), and the untested samples \( n \).
7. Randomly generate test point to determine whether it belongs to the self. If it belongs to the self,
regenerate the test point; If not, the number of test points generated by the statistics is \( N = N + 1 \);

(8) Judging whether the test point is covered by the "non-self" space, that is, whether the test point matches the existing abnormality detectors. If it matches, the test is covered, then, \( x = x + 1 \); Otherwise, the test is not covered.

(9) Calculate \( z \) according to formula (3-5) and compare with \( z_\alpha \). If \( z \) is greater than \( z_\alpha \), it means that scheduled coverage is reaching, stop training; If \( z \) is less than \( z_\alpha \), the scheduled coverage rate is unreached, and return (2), continue training.

Start
Randomly generate anomaly detector
Calculate the minimum distance \( d_{\text{min}} \) between the anomaly detector and self-sample \( S \)
Anomaly detector is enough, stop training
No
Yes
Candidate anomaly detector detector radius \( L = \delta \alpha \) n
No
Yes
Add a candidate anomaly detector to the mature detector set
No
Yes
Enter the right (2-b) process and check the coverage rate
Anomaly detector is enough, stop training
No
Yes

\[ \begin{align*}
\text{Figure 2. Anomaly detector generating process}
\end{align*} \]

3.3.3 Abnormal Detection

Next, the mature detector set generated in the previous section is used for abnormal detection and reasoning of SBS. The steps are as follows:

1) Select the untested data from SBS and perform data pre-processing.
2) Select an anomaly detector with radius \( \epsilon \) from the SBS mature anomaly detector set.
3) Calculate the distance \( D_i \) between the test data and the anomaly detector. If \( D_i < \epsilon \), the test data matched the anomaly detector. It indicated the SBS had an abnormality; If \( D_i > \epsilon \), enters step (4).
4) Determine if it is the last detector. If not, return step (2); If yes, the SBS has no abnormality at this time, then supervise the next state and return to step (1).

3.4 Fault Determination

Finally, perform fault determination.

\[ \begin{align*}
\text{Figure 3. SBS fault determination}
\end{align*} \]
(1) Detect an abnormality, find the raw data sequence, and start the expert system;
(2) Enter the inference engine to perform model matching. If yes, it means the abnormality is known, triggering the competition resolution, and selecting suitable solution; if no, it means the abnormality is unknown, then reporting the fault management module for fault location.
(3) Add the resolved fault case to SBS expert knowledge base in the specified format for the next fault detection.

4. Conclusion
The NSAEFD combines with the negative selection algorithm and the expert system proposed in this paper would bring the following benefits to SBS:
(1) The NSAEFD could detect the potential abnormal situation of the SBS in time and report it to the fault management system, which can reduce the probability of fault occurrence, improve the reliability of the communication network, and improve the user experience.
(2) The negative selection algorithm only needs to provide normal operating parameter samples when training the abnormality detection model, and it is easy to implement without providing a large number of fault samples.
(3) This mechanism combines the expert system. Each newly discovered abnormal case can be added in the expert system. When the same abnormality occurs next time, it can be processed in time to improve the stability of SBS.

Acknowledgments
This work was financially supported by National Science and Technology Major Project of the Ministry of Science and Technology of China(NO. 2018ZX03001017).

References
[1] Qian M, Wang Y, Zhou Y, et al. A super base station based centralized network architecture for 5G mobile communication systems[J]. Digital Communications & Networks, 2015, 1(2): 152-159.
[2] Y. Zhou, H. Liu, Z. Pan, L. Tian and J.L. Shi, “Cooperative Multicast with Location Aware Distributed Mobile Relay Selection: Performance Analysis and Optimized Design”, IEEE Trans. Vehicular Technology, vol. 66, issue 9, pp. 8291-8302, Mar. 2017.
[3] Zhai G, Tian L, Zhou Y, et al. Load diversity based processing resource allocation for super base stations in large-scale centralized radio access networks[C]/IEEE International Conference on Communications, 2014: 1-12.
[4] J. Wu and X. M. Li, “Communication Network Alarm Correlation Based on Multi-dimensional Fuzzy Association Rules Mining”, Second International Conference on Electric Information and Control Engineering(ICEICE), vol. 01, pp. 439-443, Apr. 2012.
[5] L. Yu, Y. Sun, K. J. Li, and M. Xu, “An Improved Genetic Algorithm Based on Fuzzy Inference Theory and Its Application in Distribution Network Fault Location”, 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA), Jun. 2016.
[6] Huang M. The Application of Artificial Immune Algorithm in Communication Network Fault Diagnosis[D]. Huazhong University of Science & Technology.2016
[7] Y. Lin, S. Ding, Y. Wang, et al, "A method of satellite network fault synthetic diagnosis based on C4.5 algorithm and expert knowledge database," 2015 International Conference on Wireless Communications & Signal Processing (WCSP), Nanjing, 2015, pp. 1-5.
[8] X. Yue, D. Wen, H. Ma, "Fault Detection Based on Real-Value Negative Selection Algorithm of Artificial Immune System," 2010 International Conference on Intelligent Computing and Cognitive Informatics, Kuala Lumpur, 2010, pp. 243-246.
[9] Z. Ji and D. Dasgupta. V-detector: An efficient negative selection algorithm with probably adequate detector coverage,"Inf. Sci. Inf. Comput. Sci. Intell. Syst. Appl: An Int. J. vol. 179, mo. 10,pp.1390-1406, Apr.2009.
[10] Chen W, Li T, Parameter analysis of negative selection algorithm, Information Sciences, Volume 420,2017,Pages 218-234.