International Journal of Applied Mathematics and Soft Computing

Volume 7, Issue 2, August, 2021

Contents

Intelligent Examination Method of Criminal Evidence based on Association Rules

Zhi Xu

School of Public Security, Northwest University of Political Science and Law, Xi'an, 710063, China

Abstract: Conventional review methods have the problem of unclear data classification, which leads to the weak relevance of data sets. This paper designs an intelligent review method of criminal evidence based on association rules. Extract the logical features of criminal evidence, combined with the interaction between data information, construct the evidence text classification model based on association rules, calculate the data mean value in the super rectangle, cluster the evidence data set information, and set the intelligent review mode according to the hidden vector feature label. Experimental results: the average correlation degree between the intelligent examination method and the two conventional examination methods is 31.425, 23.738 and 22.197 respectively, which proves that the intelligent examination method of criminal evidence combined with association rules has better performance.

Keywords: Association rules; Criminal evidence; Review method; Data set

1. Introduction

From the perspective of evidence science and academic research, the prerequisite for evidence to become the basis of verdict is: firstly, evidence must have the ability of evidence; secondly, the evidence should be proved. Only when the evidence has the ability of proof can it have the preliminary qualification to enter the court. Then, through logic and rules of experience, we can determine the value of evidence in the treatment of criminal evidence, and there is a progressive relationship between the ability of proof and the power of proof. On the contrary, the data that do not have the ability of proof lose the qualification of evidence, let alone the role of proving the facts. The method of evidence review distinguishes the ability of evidence and the power of proof, takes the ability of evidence as the entry condition of evidence into the litigation, and effectively eliminates the evidence that has a negative effect on the power of proof, so as to make the evidence review more efficient and targeted, and improve the efficiency of using evidence in litigation. All along, the traditional theory of evidence review is based on the theory of evidence attribute, which is the core of evidence application. The evidence can be used as the basis of verdict because it can meet the standards of legality, authenticity and objectivity. For example, in the process of court investigation and court debate, the judge conducts evidence investigation with the three characteristics of evidence, and the defender also conducts cross examination by explaining, questioning and refuting the three characteristics of evidence [1-3]. Just because of the advantages of simplicity and strong operability, it has gradually become the basic method of the whole criminal justice evidence review and analysis to examine the evi-

dence with three attributes and finally judge whether the evidence can be accepted. The problem is that objectivity and relevance are the content of evidence, legitimacy is the form of evidence, that is, legitimacy is the legal guarantee for the authenticity and relevance of evidence. The thinking mode of the three attributes of criminal evidence and the relationship between the three attributes are not progressive or compliant. The primacy of legality can not be revealed, and there are many risks in judicial practice [4-7].

The criminal evidence review system is a progressive process, from the evidence data to the evidence method, and then to the basis of the verdict. From the horizontal level, the evidence review method completes the procedural review and substantive review of evidence as a whole. Procedural examination refers to the examination of evidence ability. Substantive examination starts from the content of evidence itself and aims to judge the relevance and credibility of evidence. The separation of evidence access and evidence evaluation is a common feature of the evidence review system of the two legal systems. In the examination, it is necessary to change the traditional examination method of three attributes of evidence, and change it to the examination according to the logical progressive and sequential relationship between the proof ability and the proof force, which is more reasonable and in line with the law of evidence operation. After the review, the court will continue to organize the prosecution and defense to cross examine the evidence with the ability of evidence, and further judge the relationship between the evidence and the facts [7-9]. The judge will review the relevance and objectivity of the evidence according to the principle of free evaluation of

evidence, and make the judgment of the probative force. At present, the academic literature on the intelligent review method of criminal evidence is not very comprehensive and needs to be further explored.

2. Intelligent Examination Method of Criminal Evidence based on Association Rules

2.1. Extracting the logical features of criminal evidence

In essence, criminal evidence is in the form of examination and judgment of evidence, and the subject of examination and judgment is the judge trying the case, and not every criminal evidence is needed [10]. It can only be carried out when the judge has doubts about the criminal evidence rather than the opposite party, which is actually the embodiment of the function of the judge's authority. Of course, the onus probandi bears the obligation to prove that the evidence he gives is true. Of course, he needs to provide other basic evidence to prove that the evidence he gives is what he claims, so as to make the evidence admissible. In terms of content, the two evidence regulations list seven types of criminal evidence, and also stipulate the contents of examination, including: whether the media storing criminal evidence is submitted together with the printed copy, and whether there are justifiable reasons for not submitting the storage media; Whether it indicates the accessory information generated by criminal evidence and the status of relevant electronic equipment; Whether there is signature or seal of relevant personnel; whether there is forgery or alteration, etc. At the same time, it also requires that the examination and judgment of criminal evidence should be combined with other evidence to examine its authenticity and relevance. To review all kinds of criminal evidence based on judicial practice, we should not only pay attention to its source and chain of custody, but also pay attention to its authenticity and relevance, and make a comprehensive judgment combined with other evidence. China's criminal legal norms require the review of criminal evidence from acquisition to custody to presentation, while ensuring the integrity of custody chain, and clearly stipulates the identification and testing methods of authenticity, and proposes that the review of the authenticity of criminal evidence can be carried out through other verified evidence, but does not list specific methods to provide operational guidance. It relies on the examination and judgment of the recorded evidence, but does not emphasize the requirement of the relevant witness to appear in court to explain the problem, and lacks the similar provisions of self identification or presumption of criminal evidence. Although there is a pre-trial meeting in China, the purpose of the pre-trial meeting is to understand the relevant procedural issues, to hear opinions on whether there is objection to the evidence materials, and not to deal with

the objection to the evidence. Therefore, the pretrial meeting does not assume the function of judging truth. Intelligent letter review method involves many research fields of computer, and the main implementation method is information recognition. That is, how to transform the content of prisoners' handwriting into information that can be recognized by computer [11-13]. Pattern recognition refers to the process of processing and analyzing various forms of information (numerical, literal and logical) representing things or phenomena to describe, differentiate, classify and explain things or phenomena. It is an important part of information science and artificial intelligence. It can recognize characters, letters, formulas and other information. Each person's handwriting is different, so the recognition effect is different.

2.2. Construction of evidence text classification model based on association rules

In order to construct the evidence text classification model based on association rules, it is necessary to extract and classify the Chinese semantics in criminal evidence. For Chinese text, because there is no obvious segmentation mark between words, it is necessary to segment Chinese text first. Although there are many kinds of word segmentation methods, there are no more than two: one is mechanical word segmentation, which is based on word segmentation dictionary, and completes word segmentation by matching the Chinese character string in the document and the words in the vocabulary one by one. The other is understanding word segmentation, which uses Chinese grammar knowledge, semantic knowledge and psychological knowledge for word segmentation, and needs to establish word segmentation database, knowledge base and reasoning base [14, 15]. If we classify directly on this basis, it will cause a great burden on classification, so we must reduce the dimension of feature space before classification, and retain those features that contribute the most to classification. Some words such as "de" and "very" will be omitted because their contribution to text classification is very small. The so-called feature selection is to select a proper subset from the feature set, where is the size of the original feature set and the size of the selected feature set. The selection criterion is that feature selection can effectively improve the accuracy of text. Selection does not change the quality of the original feature space, but selects some important features from the original feature space to form a new low dimensional space. The calculation formula of the probability that the criminal evidence sample belongs to class C is as follows:

$$l(c|v) = \frac{l(c) + l(v|c)}{l(v)} \qquad (1)$$

In formula (1), l is the probability, c is the given document, and v is the eigenvector. When there are associa-

tion rules between feature vectors and evidence samples, formula (1) becomes the following form:

$$l\left(v\middle|c\right)=\prod_{i=1}^{n}l\left(v_i\middle|c\right) \qquad (2)$$

In formula (2), n is the number of samples i is the category of category variables, and the meaning of other variables is the same as formula (1). In the criminal proof rules, there are many rules to adjust the ability of evidence, such as the rule of excluding illegal evidence, the rule of best evidence, the rule of hearsay evidence and so on. But it needs to be clear that there is a prerequisite for the application of these rules, that is, it can be carried out on the premise that the evidence is verified. That is to say, it is the premise that other rules of evidence can be applied. Without this premise, other rules of evidence are useless. In criminal proceedings, the evidencer must verify the evidence; otherwise the electronic evidence does not have admissibility. Of course, passing the requirements of the rules of evidence test does not mean that he has the ability of evidence, but also needs to accept the test of the above rules of proof. The interaction between data and information is more complex, and the direct correlation between data is more hidden. Association analysis technology in data mining can find the relationship between different data items, and can find some implicit, even common sense can not understand the relationship. At the micro level, big data can only stay at the level of data correlation. At the macro level, the causes of the objective phenomena reflected by the data need to be explained by people's rational ability. Just as the sales of beer and diapers increase at the same time, the two sales indicators are linked because of the sales time, but they can not explain the reasons for the increase of sales at the same time, With the help of human observation and logical analysis, it is necessary to find out the causal relationship between American fathers who buy diapers for their children and buy beer for themselves. The former is a data implicated relationship, which is a weak correlation, while the latter is a causal relationship, which is "hard correlation". The relevance of the evidence used to determine the facts of the case emphasizes the life experience and rational logic of the legal person, which must have causality. Big data evidence that can not explain causality is lack of proof. In addition to the relevance of data, it also needs to meet the relevance of data and the case, which can be used as evidence to determine the basic facts of the case. The big data analysis report formed by mining massive electronic data is generally one or several documents, which is more operable to use [16]. The process of data analysis is a correlation research that expects the causation and causality between events. The correlation requirement of traditional evidence theory is higher than that of big data analysis. Therefore, it is necessary to examine the data in criminal

evidence, and the content of proof should conform to the causality of human experience logic.

2.3. Clustering evidence data set information

In the process of evidence review, it is very necessary to summarize and process the information of criminal evidence data set. The function of clustering is to divide an object set into different small object sets according to certain evaluation criteria (similar interests), and the objects in each small set have similar attributes. After clustering, the data in the same cluster has strong similarity, while the data not in the same cluster has small similarity [17, 18]. Clustering is also a kind of classification method in essence. When clustering analysis, we don't need to care about what each cluster is after clustering, just need to divide similar data into the same cluster. Therefore, for a clustering algorithm, one of its key points is to select the appropriate similarity calculation method according to the actual situation to calculate the similarity between objects. In some criminal trial practice, due to the lack of understanding of the evidence rules of data relevance, different judges have different opinions on the relevance of the same data information, and the scale of judicial practice is inconsistent. Public power organs often have a dominant position in big data forensics technology and forensics ability, It will have a completely different attitude to the relevance of big data, either hold a contradictory psychology and find that the evidence of big data is not related to the case, or ignore and avoid the relevance of big data and the facts of the case. In the subsequent continuous development process, the judicial scale is also changing within the scope of the principle. In the data set with m-dimension attribute, the mean value of data in the super rectangle is calculated as follows:

$$k = \frac{\sum_{i=1}^{t}u}{q} \qquad (3)$$

In formula (3), q is the degree of data density, t is the amount of data, u is the type of data classification, and i is the centroid of data. Once criminal evidence is collected, it needs to be preserved before it can be used in criminal trials. Preservation is a process means to maintain and protect the integrity and original state of potential electronic evidence, which means that big data needs to be transmitted and stored in a safe way to prevent tampering, and the chain of custody needs to be recorded to prove that the operation process itself does not pollute big data, In order to prevent the solidified data from being damaged or leaked, it is also necessary to restrict the authorized personnel to access the evidence. Through the formal review of the evidence collection process, we can infer the legitimacy of the evidence collection process of big data. As for the preservation of data and information, the principle of our country is to seize the storage me-

dium at the same time, with the exception of only preserving the data which cannot be seized. It is stipulated that if the original storage medium cannot be seized and the data information cannot be extracted, if the electronic data self destruction function or device exists, and the relevant data needs to be displayed and viewed on site, the relevant evidence can be fixed by printing, photographing or video recording. If the original storage medium can be seized, the original storage medium shall be seized; where the original storage medium cannot be seized but electronic data can be extracted, the data information shall be extracted. In order to protect the integrity of the data, a series of regulations in China have made provisions on the original storage medium of the impounded and sealed data. The electronic data in criminal evidence is produced by certain computer software and hardware, so to a great extent, its authenticity is closely related to the computer system. As far as the electronic data, which is completely generated by the computer system and without the intervention of artificial factors, is concerned, if the computer, corresponding software and hardware operate normally, its authenticity can be regarded as guaranteed. For the electronic data that can only be generated by manual input of certain information, the influence of computer system, software and hardware on its authenticity is also obvious. For this kind of evidence, if there is no doubt about the authenticity of the manually entered part, or if there is doubt, its authenticity can also be proved, then the normal operation of the computer, software and hardware can also guarantee its authenticity [19]. Only in a few critical moments that may have a significant impact on the authenticity of criminal data, it can operate normally. It includes the links of storage, safekeeping, analysis, transformation and presentation. Therefore, as long as the computer system attached to the criminal data runs normally in these links, unless there is evidence to the contrary, it can be presumed that it is completely generated by the computer system or even if manual input is required, However, there is no doubt about the authenticity of manual input or the authenticity of electronic data that can be reasonably interpreted.

2.4. Setting up intelligent review mode

On the basis of the criminal evidence meeting the legal provisions, the intelligent review mode should be set up. The objectivity, relevance and legality of criminal evidence are unified with each other, and they are interdependent and infiltrating each other. Electronic evidence is no exception. In order to examine and judge criminal evidence, we must start from the "three characteristics" of evidence. In order to study the rules and methods of examination and judgment of electronic evidence, we must also make clear the construction of rules and the application standards of methods. Objectivity reflects the

close combination of subjective and objective factors in criminal procedure; Relevance refers to the objective relevance between the facts of a case; Legitimacy is built on the basis of objectivity and relevance, but also contains the objectivity and relevance of evidence, which is its important guarantee. As far as the examination and judgment of criminal electronic evidence is concerned, these three attributes must be possessed at the same time. If the authenticity of physical evidence is to be accepted by the judge in court, it must be required to prove the authenticity of criminal evidence. To ensure the authenticity of electronic evidence, investigators not only need to have professional computer theoretical knowledge, but also need to rely on professional machinery and equipment and follow certain technical standards. However, in real life, many investigators will make mistakes in the process of collecting and extracting electronic evidence, because most of these judicial personnel do not have professional computer basic knowledge. Whether to follow the relevant technical standards in the process of evidence collection also needs to attract the attention of judicial personnel. As for the technical rules of criminal evidence, judicial personnel must improve their computer professional knowledge and skills, and improve their computer operation ability. Relying on advanced science and technology, we should constantly improve the intelligent examination mode of criminal evidence. Taking the text hidden vector of criminal evidence as label, the specific expression formula is as follows:

$$P = \left[p_1 \oplus p_2 \oplus \mathbf{K} \oplus p_n \right] \quad (4)$$

In formula (4), p represents the text label in the whole criminal evidence. On the basis of formula (4), the accumulation point of criminal evidence is calculated:

$$R_j^e = W^e \times G_j \quad (5)$$

In formula (5), W is the text matching label, G is the level of attention, e is the sentence level attention, and j is the hidden vector of criminal evidence. Due to the characteristics of criminal evidence itself, in the process of setting up intelligent review, it will be affected by some factors. First, because of the fragility of electronic evidence itself, it is easy to be modified, tampered or even damaged, and it is easy to be questioned by the parties. Second, because of the dependence of criminal evidence, it must be attached to a certain computer equipment to produce and transmit, so the relevant records of electronic evidence generated by computer or other carriers are easy to be questioned by the parties. Third, under special circumstances, the generation of electronic evidence in criminal evidence is also stored by a specific person or person with a specific identity, and the parties will also question its authenticity. The best evidence rule requires that in addition to the relevant legal provisions, the original documents, records or pictures must be produced. According to the corresponding technical rules,

the collected electronic evidence needs to be further checked and analyzed by means of recovery, analysis and decryption. In most cases, it also needs a series of norma-tive reviews such as technical appraisal. Based on this, complete the steps of setting intelligent review mode.

3. Experimental Analysis

In this paper, an intelligent criminal evidence review method based on association rules is designed. In order to verify the practical application performance of the design method, an experimental test is carried out.

3.1. Setting dataset metrics

On the basis of the experimental needs, the data set indi-cators are set as shown in Table 1:

elements	dence	evidence	ber
Training set	3360	2060	638
Validation set	417	247	81
Test set	409	256	79
Average duration	5.33	27.71	3.78
Average word length	11.09	3.84	7.03

In order to test the effectiveness of the design intelligent review method, the experiment selects two conventional review methods for experimental comparison, and ob-tains the experimental results.

3.2. Experimental result

The results show that the higher the value is, the better the application performance is. The experimental results are shown in Figures 1-3:

Table 1 Data set indicators

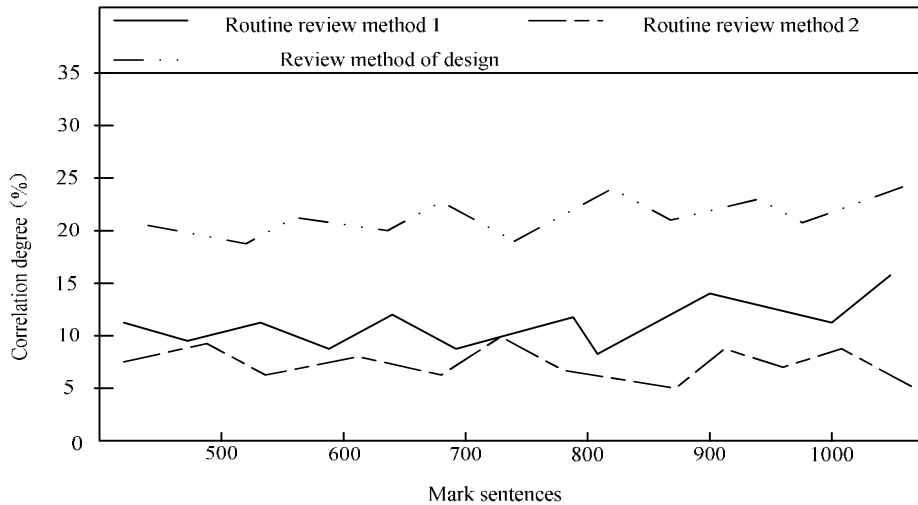| Indicators /	Name of evi-	Content of	File num-

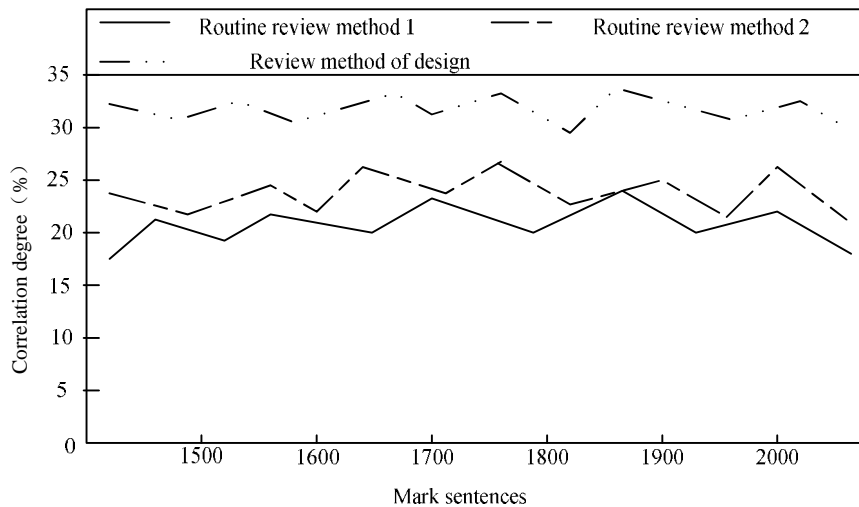**Figure 1. Annotates 1000 sentences**

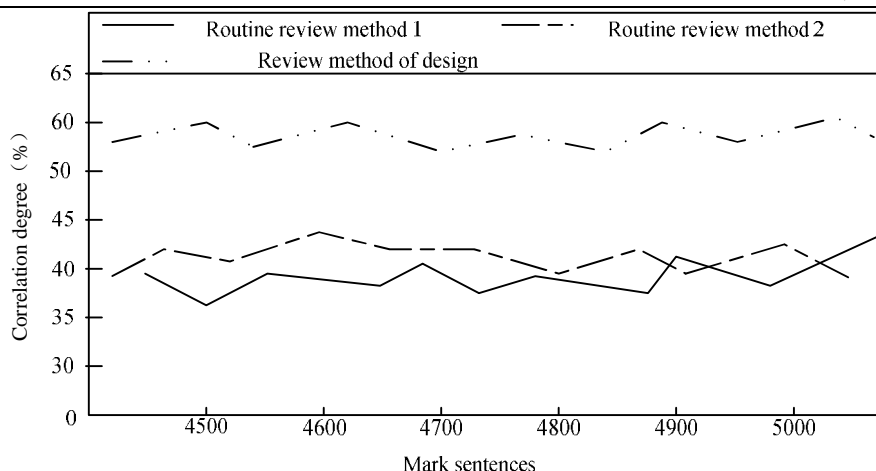

Figure 2. Annotates 2000 sentences

Figure 3. Annotates 5000 sentences l

As can be seen from Figure 1, the average correlation degree between the two conventional review methods and the intelligent review method is 11.336%, 10.207% and 20.158% respectively; As can be seen from Figure 2, the average correlation degree between the two conventional review methods and the intelligent review method is 21.098%, 23.271% and 33.112% respectively; It can be seen from Figure 3 that the average correlation degree between the two conventional review methods and the intelligent review method of this design is 38.779%, 41.006% and 54.168%, respectively, which proves that the intelligent review method of this design has better effect.

4. Conclusion

The intelligent review method designed in this paper has better performance than the conventional review method, and promotes the development process of the whole evidence review process to a certain extent. At the same time, it enriches the academic literature on criminal evidence review, and widens the application scope of association rules. Due to my limited ability, the application research of association rules in other fields is not thorough enough, and will continue to improve in the future.

References

[1] Crozier W.E., Kukucka J., Garrett B.L. Juror appraisals of forensic evidence: effects of blind proficiency and cross-examination. Forensic Science International. 2020, 315(1), 110433.

[2] Moslehi F., Haeri A., Martínez-Lvarez F. A novel hybrid GA–PSO framework for mining quantitative association rules. Soft Computing. 2020, 24(6), 4645–4666.

[3] Yu W. Discovering frequent movement paths from taxi trajectory data using spatially embedded networks and association rules. IEEE Transactions on Intelligent Transportation Systems. 2019, 20(3), 855-866.

[4] Mai T., Nguyen L., Vo B., et al. Efficient algorithm for mining non-redundant high-utility association rules. Sensors. 2020, 20(4), 1078.

[5] Aa A., Ii A., Gp A., et al. Sensing the web for induction of association rules and their composition through ensemble techniques. Procedia Computer Science. 2020, 169, 851-859.

[6] Combi C., Rizzi R., Sala P. Checking sets of pure evolving association rules. Fundamenta Informaticae. 2021, 178(4), 283-313.