From Data to Evidence: Multi-Source Evidence Association Model Based on Bayesian Network

Li Zhang¹, Feng Xu², Ming Liu²*, Tong Cui³

¹School of Public Administration & Communication, Beijing Information Science and Technology University, Beijing, China

²National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing, China

³China Judicial Big Data Research Institute, Beijing, China

*liuming@nlsde.buaa.edu.cn

Abstract. Recent years have witnessed the boom of artificial intelligence. With the advancement of judicial information, using artificial intelligence (AI) technologies to mine judicial big data is of great significance in smart courts. However, the reasoning of the evidence chain mainly relies on the judge’s manual work in the litigation process. How to model the multi-source evidence association (MSEA) and reason credible evidence chains (CEC) automatically is largely unexplored. In this paper, we propose an MSEA model based on the Bayesian network. Firstly, we construct an MSEA network in which each evidence element serves as a node, and the node correlation probability is calculated via the association relationship among the evidence elements. Subsequently, with the guidance of the event judgment chain, the MSEA model is constructed based on Bayesian networks. In the end, we use a genetic algorithm to optimize the Bayesian network and select credible evidence chains. To the best of our knowledge, this is the first time that using a probability graph model to mining the association of multi-source evidence. Experiments and the case analysis prove the effectiveness of our method.

Keywords: Bayesian network, evidence association model, judicial big data, evidence chain
1. Introduction
With the development of the rule in society, court cases have exploded largely, and the contradiction of "more cases while fewer judges" becomes the main challenge. Traditional ways have been difficult to adapt to these demands. The advance of technologies such as AI and big data may let smart courts assist judicial trials, and judicial management automatically. In judicial trials, the evidence chain is the core element of conviction in judicial trials. Only with a complete evidence chain can we provide solid support for trials and sentencing.

The judgment of a credible evidence chain is complicated, and multiple kinds of evidence often encounter conflicts of evidence or lack of key evidence. The traditional methods require judges and inspectors to analyze the contradictory and mutually corroborating of the evidence, verify the credibility and reliability of evidence, and form the complete CEC finally. The above work needs to be done manually.

The development of AI technologies such as natural language processing (NLP), knowledge graphs (KG) provides the opportunity for automatic reasoning in the evidence chain. CEC can be formed from the multiple sources of judicial evidence and the attributes of the evidence automatically, which reduce the workload of case handling effectively. Meanwhile, automatic CEC reasoning also avoids the personnel's emotional and subjective influence of judicial during the trial of cases. How to modeling MSEA and reason about the credible relationship between the evidence is the key issue in litigation analysis. As far as we know, the automatic reasoning of the evidence chain is still an open problem.

Currently, we find no method that can mine, infer, and analyze the evidence chain automatically. In most judicial cases, the plaintiff, the defendant, and the third party often provide evidence in their own interests, which are often contradicted and contradictory. This in itself challenges the credibility of the evidence. At the same time, it is also a key obstacle to the selection of credible evidence and the mining of the evidence chain. The automatic construction of the evidence chain also lacks the necessary judicial knowledge guidance. Although some scholars have applied the graph model to evidential reasoning in the judicial field [1-10], these methods have not solved the key problem of chain reasoning for credible evidence in litigation cases.

This paper constructs the MSEA model based on the Bayesian network for the first time and uses the Bayesian network to solve the problem of association modeling among multi-source evidence. First, we give out the fact-judgment chain based on the knowledge of the judicial field. These fact-judgment chains can guide the automatic construction of the evidence chain. Subsequently, based on the guidance of the event judgment chain, an MSEA network was constructed. In the MSEA, each evidence element serves as a node in the network. Based on the association relationship of the evidence elements, this characterizes the correlation probability of nodes in the network. Finally, an MSEA model based on a Bayesian network is constructed. The genetic algorithm is used to reason about the Bayesian network, and a chain of credible evidence is obtained. Experiments and case analysis of real data sets prove the effectiveness of this method.

2. Bayesian Network
The Bayesian formula is the principle. If \( P(A) \), \( P(B) \) Represents the probability of occurrence of event \( A \) and event \( B \), \( P(A|B) \) represents the probability of occurrence of \( A \) when \( B \) occurs, and \( P(A, B) \) represents the probability of occurrence of event \( A \) and \( B \) at the same time, then:

\[
P(B|A) = \frac{P(A, B)}{P(A)}
\]  

(1)

Using the above formula, we can further obtain:

\[
P(A, B) = P(A) \frac{P(B|A)}{P(B)}
\]  

(2)

The Bayesian network, also known as the belief network or the directed acyclic graphical model, is a probabilistic graphical model. After each node has the value of its direct predecessor node-set, this node condition is independent of all its indirect predecessor nodes. Its nature is very similar to the Markov process. Generally speaking, first, the multivariate non-independent joint conditional probability distribution has the following formula:

\[
P(x_1, x_2, ..., x_n) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2) ... P(x_n|x_1, x_2, ...x_{n-1})
\]  

(3)

In Bayesian networks, due to the aforementioned properties, the joint conditional probability distribution of any combination of random variables is simplified to:

\[
P(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} P(x_i|Parents(x_i))
\]  

(4)

Parents represent the union of the direct predecessor nodes of \( x_i \), and the probability value can be found from the corresponding state transition matrix. Bayesian networks are more complex than naive Bayes, and it is difficult to construct and train a Bayesian network in practice.

3. MSEA model

The construction framework of the MSEA model is shown as follows. Subsequently, the multi-source evidence network is constructed based on the fact judgment chain, and the evidence source. By matching the evidence elements with the entries in the legal knowledge rule database, a multi-source evidence fusion association map is formed based on the multi-source and heterogeneous evidence information. Finally, the genetic network is used to generate the mutated genes after the cross operation of the evidence element populations and finally optimize the algorithm.

**Fact determination chain:** The fact determination chain for different types of cases is determined based on expert experience. In-depth analysis of the litigation support chain for different types of litigation requests such as civil, criminal, and administrative. Fig.1 shows an example of the fact judgment chain for civil cases. The fact judgment chain can effectively support litigation requests. Various types of fact judgment chains constitute the judicial knowledge base.
Fact Determination chain

Fig. 1 An instance of a fact determination chain

For example, in the above instance, if there is a fact such as "battering-causing disability-causing medical expenses-incapacity-causing mental trauma", the claim for compensation can be supported. The determination of facts needs the support of evidence. In this article, the fact judgment chain is used to guide the reasoning of the evidence chain.

Evidence-related probability: Use massive historical judicial documents to extract evidence elements from them. According to the principle of maximum co-occurrence, the correlation of different types of evidence is automatically learned, and the transfer weight of multi-party and multi-type evidence is calculated. The transition probability of evidence A to evidence B is:

\[ P(A \rightarrow B) = P(B|A) = \frac{p(B|A)}{p(B)} \] (5)

In this paper, the transition probability of evidence entities is considered as the node weight in MSEA. After selecting the corresponding fact-judgment chain template, we classify the evidence according to the plaintiff's evidence, the defendant's evidence, the forensic evidence, and the third-party evidence.

Fig. 2 Schematic diagram of the construction of evidence network nodes
MSEA model and reasoning: Based on the MSEA model, we use optimization methods such as a genetic algorithm to reason about the MSEA network and calculate all possible combinations of evidence chains. We select the evidence chain with the largest probability serves as the CEC, which is shown in the following figure.

![Evidence Network](image)

**Fig. 3** Evidence chain reasoning based on MSEA model

4. Experiment and case analysis

The method is implemented using Java and Matlab mixed programming. The data set comes from the real-world data crawled by the referee document website. We manually classified and screened out three typical cases: Loan, Road Traffic, and Criminal cases. Meanwhile, judicial experts in our group labeled the evidence chains of 10 litigation cases in each type of case; then, two groups of experts cross-checked the evidence chains marked by each other. These labeled evidence chains can be used as ground truth to test our MSEA model.

Jaccard similarity: We use Jaccard similarity to measure the correlation of the automatically-mined evidence chain and the real evidence chain of the case. The larger the Jaccard coefficient, the higher the degree of overlap between the real evidence chain and the predicted CEC.

| Chain Number | Criminal | Road traffic | Loan |
|--------------|----------|--------------|------|
| 1            | 0.904    | 0.347        | 1.00 |
| 2            | 0.870    | 0.5393       | 1.00 |
| 3            | 0.816    | 0.565        | 0.988|
| 4            | 0.757    | 0.583        | 0.963|
| 5            | 0.689    | 0.607        | 0.915|
| 6            | 0.604    | 0.347        | 0.836|
| 7            | 0.519    | 0.836        | 0.658|
| 8            | 0.451    | 0.343        | 0.381|
| 9            | 0.412    | 0.349        | 0.492|
| 10           | 0.344    | 0.454        | 0.5  |

**Tab. 1** The similarity between the predicted and the real evidence chain
**Manual evaluation:** A total of 30 evidence chains need to be predicted for the three types of litigation cases. The human experts in our research group manually verify the degree of consistency between our CEC and the real evidence chain. The accuracy rate of the evidence chain reasoning for the three types of causes is shown in the figure below:

![Fig. 4 Reasoning accuracy rate of different types of evidence chains](image)

5. **Conclusion**
To automatic reason the evidence chain, this paper proposes a Bayesian network-based MSEA model for the first time. Bayesian networks are used to model MSEA for the first time. This research was supported by the Research Foundation of Beijing Information Science and Technology University (2035015), CNCERT Key Foundation for Youths (2020Q08), the Basic Research Project of Military Commission of Science and Technology (2017–JCJQ–ZD–043–04), and the National Key Research and Development Project (2016QY04W0901).

**References**

[1] Pal K, Campbell J A 1997 An application of rule-based and case-based reasoning within a single legal knowledge-based system *J. ACM SIGMIS Database* **28**(4) pp 48-63

[2] Juisheng C 2012 Comparison of multilabel classification models to forecast project dispute resolutions *J. Expert Systems with Applications* **12** pp 10202-10211

[3] Timmer S T, Meyer J J C, Prakken H and et al 2015 A structure-guided approach to capturing Bayesian reasoning about legal evidence in argumentation *Proc.15th International Conference on Artificial Intelligence and Law* (San Diego California: ACM) pp 109-118

[4] Vlek C, Prakken H, Renooij S and et al 2015 Constructing and understanding Bayesian networks for legal evidence with scenario schemes *Proc. 15th International Conference on Artificial Intelligence and Law* (San Diego California: ACM) pp 128-137

[5] Liu M, Lang B and Gu Z 2018 Similarity calculations of academic articles using topic events and domain knowledge *Proc. Asia-Pacific Web and Web-Age Information Management Joint International Conference* (Macao: Springer) pp 45-53

[6] Curtotti M, McCreath E, Bruce T and et al 2015 Machine learning for readability of legislative sentences *Proc.15th International Conference on Artificial Intelligence and Law* (San Diego California: ACM) pp 53-62

[7] Juisheng C 2012 Comparison of multilabel classification models to forecast project dispute resolutions *J. Expert Systems with Applications* **12** pp 10202-10211
[8] Arditi D and Tokdemir O B A 1999 comparison of case-based reasoning and artificial neural networks *J. Comput. Civ. Eng* 13(3) pp 162-169

[9] Pal, S K and Shiu, S C K 2004 Foundations of soft case-based reasoning Wiley Online Library pp161-200

[10] Ronald J Allen The Relevance and Admissibility of Evidence 2010 Evidence Science No. 3

[11] Ming L, Bo L, Zepeng G and Ahmed Z 2017 Measuring similarity of academic articles with semantic profile and joint word embedding *J. Tsinghua Science and Technology* 22(6) pp 619-632

[12] Arditi D and Tokdemir O B 1999 A comparison of case-based reasoning and artificial neural networks *J. Journal of Computing in Civil Engineering* 13(3) 162-169