Intelligent Manufacturing and Cloud Computing
I.S. Jesus and K. Wang (Eds.)

doi:10.3233/ATDE241378

Application of NLP Algorithms in Compliance Management

Xiaorong Qiu[a,1]
[a]*LIANCHU SECURITIES CO.,LTD., Shenzhen 518000, Guangdong, China*

Abstract. To address the challenges posed by complex regulations and policies in compliance management, Natural Language Processing (NLP) algorithms are employed to enhance the efficiency and accuracy of text analysis. Utilizing BERT and Transformer as core models, these systems automatically generate compliance documents, categorize compliance-related texts, and dynamically monitor user behavior, significantly improving compliance detection and risk assessment. Data shows that text classification based on BERT achieved an accuracy of 93.4%, and the efficiency of automatic compliance document generation is 150 times higher compared to manual writing, with user satisfaction for automated compliance review maintaining above 85%. Analysis suggests that the application of NLP technology in compliance management can significantly reduce compliance costs and enhance risk alert capabilities.

Keywords. Natural language processing, compliance management, risk assessment

1. Introduction

The challenges in compliance management are increasingly prominent, with the complexity of regulations and policies making traditional methods insufficient. Natural Language Processing (NLP) technology, with its strengths in text comprehension, semantic analysis, and information extraction, provides a new pathway for compliance management, enhancing efficiency and accuracy through automation and intelligent solutions. Deep learning algorithms are employed to analyze complex legal texts, aiming to reduce manual costs and improve decision accuracy. It is anticipated that more efficient applications can be achieved in areas such as text classification, risk assessment, and dynamic monitoring, thereby lowering compliance risks, optimizing corporate strategies, and strengthening legal resilience and competitive advantage.

2. The Importance of Natural Language Processing in Compliance Management

Natural Language Processing (NLP) plays a crucial role in compliance management, focusing on the automated and intelligent handling of vast amounts of textual data, making compliance management more efficient and accurate. With the increasing complexity of regulations and policies, traditional compliance inspection methods struggle to manage large volumes of text data [1]. NLP technology can rapidly extract

[1] Corresponding Author: Xiaorong Qiu, qiuxiaorongsz@163.com

key information from extensive documents through semantic analysis, entity recognition, and text classification, reducing the workload of manual review. Additionally, NLP can identify potential compliance risks in a timely manner by monitoring and analyzing public information, social media, and internal data, thereby enhancing risk alert capabilities. Furthermore, NLP enables enterprises to remain agile amidst various legal and policy changes, quickly adjusting compliance strategies and improving overall compliance and decision accuracy. This technology not only enhances the efficiency of compliance management but also reduces legal risks and operational costs for companies.

3. Application Models of NLP Algorithms in Compliance Management

3.1. Overview of NLP algorithms

The application of Natural Language Processing (NLP) algorithms in compliance management relies on the collaboration of various technologies to process and analyze large volumes of text data in an automated and intelligent manner. NLP algorithms encompass several steps, including text preprocessing, semantic analysis, sentiment detection, and entity recognition, accurately interpreting legal, policy, and compliance-related information within the text. In compliance management scenarios, common application models include text classification models, Named Entity Recognition (NER) models, and sentiment analysis models. Text classification models utilize machine learning algorithms to categorize documents, identifying non-compliant content; NER models are employed to extract key information, such as legal clauses and company names, from contracts, reports, and other documents for further analysis; sentiment analysis models are particularly significant for monitoring social media and public sentiment changes, detecting potential compliance risk signals [2]. The implementation of these models not only improves the accuracy and efficiency of compliance audits but also enables quick adaptation to regulatory changes, helping companies maintain a competitive edge in complex compliance environments.

3.2. Compliance text classification based on the BERT model

Compliance text classification using the BERT (Bidirectional Encoder Representations from Transformers) model is one of the advanced applications of Natural Language Processing in compliance management [3]. BERT, with its bidirectional Transformer architecture in deep learning, can capture complex semantic relationships within the text's context, enabling precise classification of compliance-related documents. In the process of compliance text classification, the BERT model first converts text data into vector representations through tokenization and embedding layers. This step can be expressed by the following formula:

$$X = [x_1, x_2, \dots, x_n] \tag{1}$$

In this context, X represents the input text, and x_n denotes the embedding representation of each word. Next, the BERT model calculates the relationships between words in the text using the Self-Attention mechanism. This process can be expressed as:

$$Attention(Q,K,V) = soft\max\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

In this mechanism, Q, K, and V represent the query, key, and value matrices, respectively, while d_k is the dimension of the key. This mechanism allows BERT to consider the associations between all words in the text, leading to a deeper understanding of the content. In practical applications, BERT is fine-tuned using large-scale pre-trained corpora to adapt to specific compliance classification tasks. The model undergoes supervised learning with domain-specific datasets, optimizing model parameters by maximizing the log-likelihood function:

$$L = -\sum_{i=1}^{N}\left[y_i \log(p(y_i|X_i)) + (1-y_i)\log(1-p(y_i|X_i))\right] \tag{3}$$

In this context, N represents the number of training samples, y_i is the true label of the i sample, and $p(y_i|X_i)$ denotes the predicted probability by the model. Through this optimization process, BERT efficiently performs text classification, accurately identifying non-compliant content. Additionally, BERT's multi-layer encoder captures subtle distinctions within compliance texts, aiding in the recognition of complex legal terms and policy documents. This precise classification capability significantly enhances the automation level of compliance management, reducing both the cost and time of manual review [4].

3.3. Multilingual compliance handling based on transformer

The Transformer model, leveraging its Self-Attention mechanism and multi-layer encoder architecture, effectively addresses the complexity of multilingual semantic understanding, providing strong support for compliance processing [5]. Due to differences in grammatical structures and semantic expressions across languages, traditional methods often face challenges in multilingual scenarios. In contrast, the Transformer excels in parallel processing and cross-linguistic information integration, making it highly effective in handling multilingual compliance texts. The core of the Transformer in multilingual compliance handling is the Multi-Head Attention mechanism, which allows the model to simultaneously focus on different parts of the text from various semantic dimensions. The computation process of Multi-Head Attention can be represented as:

$$MultiHead(Q,K,V) = Concat(head_1, head_2, \ldots, head_n)W^O \tag{4}$$

In this context, $head_h = Attention(QW_i^Q, KW_i^K, VW_i^V)$, W_i^Q, W_i^K, and W_i^V are the weight matrices used for linear transformations, with W^O being the output weight matrix. This mechanism enables the Transformer to extract multi-level semantic features simultaneously from texts in different languages, allowing for accurate analysis of compliance-related content. In multilingual processing scenarios, the Transformer often employs a shared vocabulary strategy, utilizing Subword Embedding to address the issue of vocabulary misalignment between languages. This process can be represented by the following formula:

$$Embedding(x_i) = \sum_{j=1}^{m} E_j \cdot Subword(x_i)_j \tag{5}$$

In this context, x_i represents the input word or phrase, $Subword(x_i)_j$ denotes the j subword segment, and E_j is the corresponding embedding vector. Through this method, the Transformer can map multilingual texts into a unified high-dimensional semantic space, achieving semantic alignment across different languages. In practical applications, multilingual compliance processing models based on Transformer can be fine-tuned for specific compliance tasks [6]. Leveraging pre-training and transfer learning, the model can quickly adapt to various linguistic environments and industry contexts, achieving high-precision compliance detection and text classification.

4. Application of NLP in Compliance Management

4.1. Automatic generation of compliance documents

Through Natural Language Generation (NLG) technology, systems can automatically analyze and integrate a vast array of regulations, policies, and compliance documents, achieving efficient and accurate document generation. The core of this technology lies in leveraging deep learning algorithms, particularly Transformer-based language models, to semantically interpret complex legal language and specialized terminology, reconstructing them into compliant reports and legal documents [7]. See Figure 1 for an analysis of accuracy.

As shown in Figure 1, the automatic document generation using NLP demonstrates a high accuracy rate across various types of compliance tasks, with a difference of less than 3% compared to manually written results, indicating its reliability in terms of accuracy. The BERT and GPT models performed the best in the task of generating compliance review reports, achieving accuracies of 93.4% and 92.1%, respectively. The next section analyzes the efficiency improvements in compliance document generation, as illustrated in Table 1.

Table 1. Efficiency comparison of compliance document generation

Document Type	Automatic Generation Time (seconds)	Manual Writing Time (minutes)	Efficiency Improvement (times)
Compliance Review Report	12	30	150
Contract Clause Summary	9	25	167
Regulation Change Notice	8	20	150
Risk Assessment Report	15	35	140

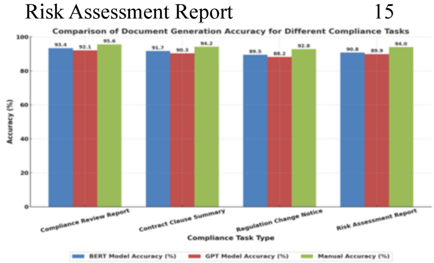

Figure 1. Accuracy comparison of document generation for different compliance tasks

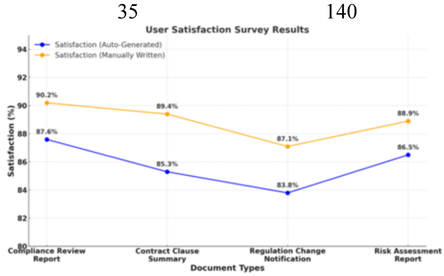

Figure 2. User satisfaction survey

Table 1 shows the significant efficiency gains achieved with NLP-based automatic compliance document generation. For instance, generating a compliance review report takes only 12 seconds, while manual writing requires 30 minutes, highlighting the substantial time-saving advantage of NLP technology. Figure 2 further explores user evaluations of different generation methods through a user satisfaction survey.

The data above compares user satisfaction between automatically generated and manually written compliance documents. Although the automatically generated documents are slightly lower in some aspects compared to manual writing, overall satisfaction remains above 85%, indicating that it effectively meets user needs in most scenarios.

4.2. Automated compliance review and risk assessment

Leveraging Natural Language Processing, particularly deep learning models, enables the automated parsing and analysis of complex regulatory documents, contract content, and relevant legal clauses [8]. This automated review process, through semantic understanding and text analysis, accurately identifies risk areas and provides clear risk assessment reports. To illustrate the effectiveness of automated compliance review and risk assessment, an analysis of the accuracy rates of different algorithms in compliance review is shown in Figure 3.

As shown in Figure 3, deep learning algorithms based on Natural Language Processing outperform traditional rule-based matching algorithms in compliance reviews. BERT-based models show a significant improvement in both precision and recall, reaching 91.2% and 88.7% respectively. In terms of time efficiency, automated compliance reviews demonstrate a considerable advantage over manual reviews. A comparison of time costs for different review methods is provided in Table 2.

Table 2. Time cost comparison for compliance review

Review Type	Manual Review Time (minutes)	Automated Review Time (seconds)	Time Efficiency Improvement (times)
Regulatory Compliance Check	40	15	160
Contract Clause Review	35	12	175
Risk Assessment Report	50	20	150

The data in Table 2 indicates that automated reviews significantly reduce the time required for compliance and risk assessments. A regulatory compliance check takes only 15 seconds, improving time efficiency by 160 times. To further evaluate the quality of automated compliance reviews, a user satisfaction survey was conducted comparing manual and automated reviews, with the results presented in Table 3.

Table 3. User satisfaction comparison between manual and automated compliance review

Review Type	Satisfaction (Manual Review)	Satisfaction (Automated Review)
Regulatory Compliance Check	92.40%	88.90%
Contract Clause Review	90.10%	87.30%
Risk Assessment Report	89.70%	86.40%

The data in Table 3 shows that although manual reviews slightly exceed automated reviews in satisfaction, automated review satisfaction consistently remains above 85%, demonstrating a high level of practicality and acceptance in compliance management. A comprehensive analysis of the data indicates that automated compliance review and risk assessment show clear advantages in accuracy, efficiency, and user feedback.

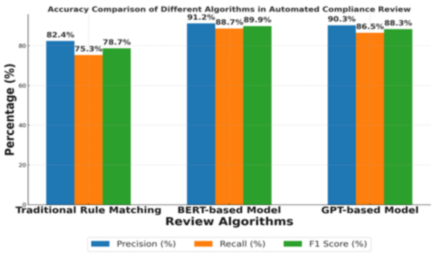

Figure 3. Accuracy comparison of different algorithms in automated compliance review

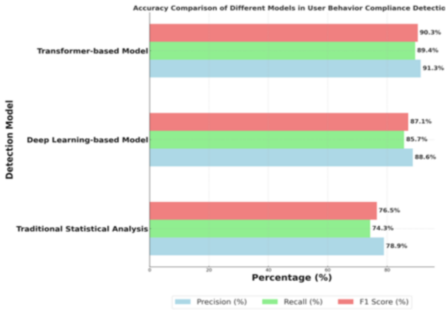

Figure 4. Accuracy comparison of different models in user behavior compliance detection

4.3. Dynamic compliance monitoring based on user behavior

Compared to traditional static compliance review methods, dynamic monitoring continuously observes changes in user behavior within the system, promptly identifying and responding to potential violations. The core of this technology lies in leveraging data such as user behavior logs, access records, and operation traces, in combination with machine learning and deep learning algorithms, to achieve real-time risk alerts and dynamic updates of compliance reviews [9]. Starting from the accuracy of compliance detection based on user behavior data, Figure 4 compares the detection accuracy of different models in dynamic compliance monitoring.

Figure 4 shows that Transformer-based models perform best in user behavior compliance detection, achieving a precision of 91.3%, significantly surpassing traditional statistical analysis models. This indicates that deep learning models have superior capability in handling complex behavioral data. Dynamic compliance monitoring also demonstrates considerable advantages over traditional methods in monitoring frequency and response time. Table 4 compares the response time and monitoring frequency of different monitoring methods.

Table 4. Comparison of response time and monitoring frequency between different compliance monitoring methods

Monitoring Method	Average Response Time (seconds)	Monitoring Frequency (times/minute)	Monitoring Coverage (%)
Static Compliance Check	180	0.1	65
Dynamic Compliance Monitoring	15	4.5	95

Table 4 highlights the significant improvements in response speed and monitoring frequency achieved by dynamic compliance monitoring. The average response time of dynamic monitoring is 15 seconds, much faster than the 180 seconds of static compliance checks. Additionally, dynamic monitoring reaches a frequency of 4.5 times per minute, with a monitoring coverage of 95%, which is markedly better than the 65% of static checks. These findings indicate that dynamic compliance monitoring based on user behavior excels in real-time detection and coverage, enabling companies to identify potential compliance risks more accurately and promptly. With the integration of machine learning and deep learning technologies, dynamic compliance monitoring not only improves the accuracy of compliance detection but also significantly reduces the response time for risk identification, enhancing the overall efficiency of compliance management [10].

5. Conclusion

The application of Natural Language Processing technology in compliance management demonstrates efficiency and accuracy, effectively enhancing compliance detection and risk alert capabilities through methods such as text classification, automatic compliance document generation, and dynamic compliance monitoring. Future efforts can focus on optimizing model adaptability to multilingual contexts and improving the precision of handling complex legal language, providing smarter and more comprehensive solutions for compliance management.

References

[1] Boning J .Analysis of the “Trinity” Model of Corporate Compliance Management[J].Journal of Electronic Research and Application,2024,8(3):7-12.

[2] Zheng D ,Xu Y ,Wenren Y .Compliance management and investment efficiency in state-owned enterprises: Evidence from China[J].China Journal of Accounting Research,2024,17(2):100358-.

[3] Yi Z .Research on Enterprise Legal Service Strategy Based on Compliance Management Optimization[J].Proceedings of Business and Economic Studies,2024,7(2):84-88.

[4] Chen X ,Hu Z ,Wang C .Empowering education development through AIGC: A systematic literature review[J].Education and Information Technologies,2024,29(13):17485-17537.

[5] Rafiee A ,Feyzi F ,Shahbahrami A .Electronic Waste Management Using Smart Contracts on the Blockchain Platform[J].SN Computer Science,2024,5(7):896-896.

[6] Hu C ,Liu Z ,Li R , et al.Smart Contract Assisted Privacy-Preserving Data Aggregation and Management Scheme for Smart Grid[J].IEEE Transactions on Dependable and Secure Computing,2024,21(4):2145-2161.

[7] Xiaohui Y ,Kun Z ,Kai J , et al.A personalized and efficient EMR sharing and management scheme based on smart contracts[J].Peer-to-Peer Networking and Applications,2024,17(3):1569-1587.

[8] Arshed A M ,Gherghina C Ș ,Dewi C , et al.Unveiling AI-Generated Financial Text: A Computational Approach Using Natural Language Processing and Generative Artificial Intelligence[J].Computation,2024,12(5):101-.

[9] Abdul W ,Jun W ,Alireza S , et al.A model-based smart contracts system via blockchain technology to reduce delays and conflicts in construction management processes[J].Engineering, Construction and Architectural Management,2023,30(10):5052-5072.

[10] J. R ,E.Q. W ,R. A , et al.MSR131 Prospects for Automation of Systemic Literature Reviews (SLRs) With Artificial Intelligence and Natural Language Processing[J].Value in Health,2023,26(12S):S418-S418.