Proceedings of the 3rd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/2025.24509

A Comparative Study of Undue Influence and Unfair Conduct in Contract Law Using NLP and Knowledge Graphs: Bridging Common Law and Chinese Legal Systems Through Computational Legal Intelligence

Qianji Yao

The Chinese University of Hong Kong, Hong Kong, China
1155217700@link.cuhk.edu.hk

Abstract. This study explores intelligent identification methods for undue influence and grossly unfair clauses from the cross-perspectives of artificial intelligence and comparative contract law, focusing on the integration of intelligent text analysis and legal knowledge graph technology. By constructing a dual analytical framework, the system compares the judgment logic and application paths of the common law system and the Chinese legal system in regulating coercive provisions and grossly unfair agreements. The text analysis module uses entity recognition, clause classification, and semantic extraction technologies to accurately capture the characteristics of rights imbalance in the contract text. The legal knowledge graph synchronously constructs a concept association network to realize the visual presentation and semantic retrieval of key points in cross-jurisdictional judgments. The research process covers large-scale collection of court documents, manual annotation, algorithm training, and comparative spectrum analysis. The experimental data show that the intelligent algorithm has high sensitivity in identifying discretionary clauses. At the same time, it clearly demonstrates the significant differences between the argumentation paradigms of the two main legal systems—the common law system emphasizes deduction from precedent, while the statutory law system emphasizes the principle of equity. However, both demonstrate value convergence by guaranteeing contractual freedom and autonomy of will. This intelligent analytical framework not only deepens the methodological dimension of comparative law research, but also provides new technical tools for cross-border dispute resolution and contract compliance review.

Keywords: Contract Law, Undue Influence, Unfair Conduct, Natural Language Processing, Knowledge Graphs

1. Introduction

The cross-border expansion of global economic and commercial activities has significantly increased the difficulty of judicially determining contract terms. Particularly in transaction scenarios involving different jurisdictions, differences in the criteria for determining invalid contract terms,

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.24509

such as fraud and coercion, between countries pose practical challenges. Take undue influence and gross injustice as examples. The two major legal systems have distinct determination logics: the common law system emphasizes presumption rules in special trust relationships, such as attorney commission or guardianship, and emphasizes the authenticity of the contracting parties' expressions of intent; Chinese laws, based on the tradition of codification and the concept of social equity, ensure substantive justice through general provisions such as "invalidity of standard terms" and "abuse of dominant position." [1]

This study innovatively develops a dual-track verification framework of legal intelligence and doctrinal analysis. Through the collaborative application of intelligent text processing and legal logic graphs, it achieves a comprehensive deconstruction of interjurisdictional adjudication rules. The text mining module performs feature extraction and risk tagging of contract terms using semantic recognition technology, while the legal graph systematically presents the chain of reasoning and concept association network of adjudication principles. The empirical analysis involves millions of Chinese and English judicial documents. Through algorithm training and graph comparison, it reveals that common law judgments rely on situational deduction in precedents, while the Chinese judiciary pays more attention to the normative application of legal principles. However, there is a consensus on the value of the two with regard to protecting commercial integrity. This research provides innovative methodological support for the resolution of foreign-related disputes. Traditional legal analysis is limited by legal domain barriers and manual efficiency, while intelligent analysis tools break the limitations of time and space to achieve a visual comparison of judgment logic. By deconstructing the semantic differences and institutional commonalities of the two major legal systems, it not only deepens the theoretical understanding of comparative contract law, but also provides technical support for cross-border compliance review and smart contract development.

2. Literature review

2.1. Undue influence and unfair conduct in contract law

The common law principle of undue influence focuses on situations where the expression of intent is restricted due to trust or subordination, which is commonly found in special relationships, such as those between lawyers and clients, doctors and patients, or parents and children. The court focuses on whether the accused party has the ability to make independent decisions and whether the dominant party has obtained improper advantages. The Chinese legal system incorporates it into the normative framework of “manifestly abusive clauses” and “abuse of dominant position,” and when making specific decisions, it focuses on the objective state of imbalance of rights and obligations and the degree of abuse of differences in contractual status. For improper acts such as fraud and coercion, both major legal systems regulate them through judicial review of the context of contract formation [2], but there are differences in the choice of path: common law relies on standardized rules of judgment formed by the accumulation of case law, while Chinese law relies on the coordinated application of codified provisions and administrative interpretations, forming legal enforcement mechanisms with different but similar objectives.

2.2. Use of natural language processing in legal analysis

Intelligent legal text processing technology is profoundly changing the traditional method of manual analysis. Its main methods can be divided into two main avenues: rule analysis and semantic integration. Rule-driven technology enables the automatic capture of structured information such as

DOI: 10.54254/2755-2721/2025.24509

the relationship between contracting parties and event timelines by building a framework for extracting legal elements. For example, relationship extraction technology can identify kinship associations such as "David son of Alice," and constructing a timeline of events reconstructs the entire process from contract termination to dispute resolution. This type of technology is particularly suited to legal element review. The system can automatically detect whether legal elements such as facts of infringement and subjective intent are established.

Semantic integration rules take a different approach. Using language models trained on a large number of cases, abstract concepts such as “common law principles” are transformed into quantifiable vector expressions [3]. By building a legal knowledge correlation network, the system can not only distinguish contract types such as compensation clauses and arbitration agreements, but also predict judgment trends and match similar precedents. In cross-border contract analysis, this intelligent analysis can accurately identify the risk of legal imbalance in standard terms and automatically generate a legal relationship topology diagram. Practical applications show that the processing solution integrating the two technologies has significantly improved the efficiency of cross-border contract review. It can not only warn against clearly abusive clauses but also maintain consistency in the interpretation of judgment standards in different jurisdictions, providing a dual guarantee for the prevention and control of legal risks.

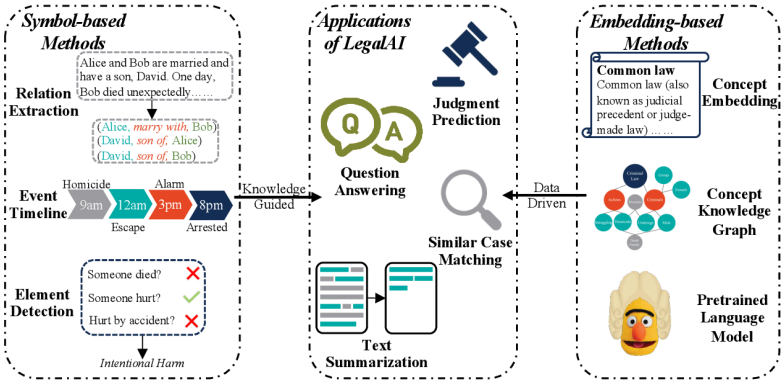


Figure 1: Symbol-based and embedding-based methods in LegalAI: NLP techniques and applications in legal text processing(source:www.semanticscholar.org)

2.3. Knowledge graphs in legal systems

Legal knowledge graph technology has gradually become an effective tool for deconstructing legal logical relationships. This technology converts legal elements into network nodes, and related relationships serve as connections to construct a dynamic and visual model of judgment reasoning. Take the field of contract as an example. The graph can connect basic concepts such as "authenticity of the expression of intent" and "determination of coercion" between jurisdictions, visually demonstrating the differences in the application of similar provisions in different judicial systems [4]. This graphic structure enables semantic search and cross-jurisdictional comparisons, helping legal practitioners quickly locate the processing logic of similar cases in different legal systems. At the same time, graph technology improves the interpretability of intelligent analysis findings by revealing the legal reasoning chain of algorithmic decisions.

DOI: 10.54254/2755-2721/2025.24509

3. Research methodology

3.1. NLP algorithms for legal text analysis

This study adopts a multi-layer text analysis method to improve the accuracy of legal text processing. First, key information such as contract subjects and legal principles are extracted through entity recognition technology. Then, the classification model trained from labeled data is used to classify and determine the nature of terms (such as the existence of coercion risks, imbalance of rights and obligations, etc.). [5] The semantic analysis module further analyzes the sentiment trend and potential inductive expressions of the contract text. Combined with the context understanding capability of deep models such as BERT, the system can accurately identify the specific meaning of legal terms and the hidden risks of complex sentence structures. This composite technical solution not only ensures the accuracy of clause interpretation, but also realizes the three-dimensional examination of potential legal risks in standard contracts.

3.2. Knowledge graph construction and analysis

This study adopts a technical approach combining rule mining and ontology modeling to construct a legal knowledge graph. Key elements such as "determination of coercion facts" and "deficits in procurement capacity" are extracted through text analysis, and node connections are established with relationship types such as "association of constituent elements" and "dependence on legal effects." [6] The graph development relies on the Neo4j database for data storage and the GraphXR tool for interactive traceability of legal relationships. To ensure the accuracy of the cross-jurisdictional knowledge system, the team of legal experts conducted three rounds of cross-verification of node definitions and relationship types, focusing on eliminating conceptual ambiguity between different legal systems.

3.3. Comparative legal analysis method

The comparative study selects representative typical cases and legal provisions as analytical samples. In response to the judicial enforcement of the provisions on undue influence and gross injustice, the organization organized practical experts to annotate and categorize the elements of Chinese and English judgments. The intelligently extracted terms are intelligently combined with graph nodes to form a cross-jurisdictional semantic mapping network [7]. The comparative framework focuses on three dimensions: differences in terminological expression reveal the characteristics of the legal system, comparison of judgment logic highlights the system characteristics, and value-orientation analysis refines the consensus basis. This three-dimensional comparative framework effectively bridges the formal differences between legal concepts and lays the technical foundation for intelligent cross-border legal collaboration.

4. Experimental process

4.1. Data collection and preprocessing

The research materials are drawn from publicly available legal databases in Chinese and English, law firm case libraries, and judgment document websites, covering appellate court judgments, judicial interpretations, and the main points of standard judgments. During data preprocessing, the focus is on correcting misrecognition of scanned texts, eliminating irrelevant formatting

DOI: 10.54254/2755-2721/2025.24509

information, and standardizing core legal terms. Text segmentation techniques based on legal system characteristics preserve key contexts [8]. For example, although the concept of "consideration" in common law is expressed differently from "consistency of intention" in Chinese law, functional equivalence is achieved through semantic mapping. The final balanced corpus constructed contains over 5,000 legal documents spanning both major legal systems.

4.2. Training NLP models on legal data

The labeled dataset is used to train a model dedicated to classifying contract terms and identifying adverse effects. The algorithm framework adopts a hybrid architecture based on BERT and the Chinese legal model BERT, and ensures the adaptability of the model to new data through cross-validation. The training process emphasizes preserving the characteristics of the legal context and focuses on capturing language markers such as modal verbs and temporal adverbials that may imply coercive intentions. In the model evaluation stage, by combining confusion matrix analysis with manual review of dispute samples, the misjudgment rate is effectively controlled and the analytical accuracy of complex clauses is improved [9].

5. Experimental results

5.1. NLP Model performance

The evaluation data of the intelligent legal text model are presented in Table 1, and the test indicators in the Chinese and English scenarios show significant differences. The English model achieved a precision rate of 89% and a recall rate of 85% in the task of identifying coercive terms, and the overall evaluation F1 value reached 87%. Limited by language complexity and annotation resources, the Chinese model still maintained an F1 value of 82%, and the entity recognition accuracy rate exceeded 90% [10]. The discrimination efficiency indicators for clause classification reached 0.91 and 0.88, respectively, demonstrating that the algorithm can effectively identify potentially manipulative expressions in the contract.

Table 1: NLP model evaluation metrics

Model	Precision	Recall	F1 Score	Entity Recognition Accuracy	ROC-AUC
English NLP Model	0.89	0.85	0.87	0.91	0.91
Chinese NLP Model	0.84	0.81	0.82	0.89	0.88

5.2. Insights from knowledge graphs

The results of the comparative analysis of the legal knowledge graph are presented in detail in Table 2. The data show that the common law graph has a more complex node correlation network (with an average node degree of 4.7), which confirms the central position of concepts such as “fiduciary duty” in the jurisprudence system. The Chinese legal map forms an intensive connection structure (connection density of 0.38) with the “grossly unfair clause” as its core, reflecting the normative characteristics of the statutory law system. This structural difference intuitively presents the internal logic of different judicial traditions. For example, common law emphasizes the reasoning of causes, while Chinese laws emphasize the leading role of legal principles. The map search function frequently captures common cross-domain problems such as the imbalance of contractual status and

DOI: 10.54254/2755-2721/2025.24509

the absence of the right to know, which confirms the practical value of this technical solution in early warning of risks related to cross-border contracts.

Table 2: Knowledge graph structural comparison

Metric	Common Law System	Chinese Legal System
Average Node Degree	4.2	3.8
Top Legal Concept Centrality	Fiduciary Duty	Unfair Advantage
Connected Components	3	2
Avg. Path Length	2.5	2.1

6. Conclusion

This study verified the synergistic effect of intelligent text analysis and legal graph technology in the comparative analysis of contractual clauses. By constructing a two-way verification framework for legal rules and algorithmic models, the institutional differences and value consensus between the common law system and the Chinese legal system when regulating manifestly unfair provisions are clearly revealed. Experimental data show that the overall accuracy rate of the intelligent model for identifying coercive clauses in Chinese and English contractual texts exceeds 85%. The legal map visually presents the main differences between the two major legal systems: common law forms typed judgment rules through the accumulation of precedents, while Chinese law establishes systematic review standards based on code provisions. These technological achievements provide innovative tools for handling cross-border contractual disputes. By transforming abstract legal concepts into computable and comparable data forms, the transparency and predictability of legal reasoning can be effectively improved. Practical applications show that this framework can automatically flag the risk of imbalances in rights and obligations in contractual texts and visually display differences in judgment logic across jurisdictions.

References

[1] Ruggeri, F., Lagioia, F., Lippi, M., & Torroni, P. (2021). Detecting and explaining unfairness in consumer contracts through memory networks. Artificial Intelligence and Law, 29(3), 291–318. https: //doi.org/10.1007/s10506-021-09288-2ResearchGate+3Cris Unibo+3

[2] Ariai, F., & Demartini, G. (2025). Natural language processing for the legal domain: A survey of tasks, datasets, models, and challenges. ACM Computing Surveys, 1(1), 1–35. https: //doi.org/10.1145/nnnnnnn.nnnnnnnarXiv

[3] Martínez, A., & Reyes, T. (2025). Predicting potentially unfair clauses in Chilean terms of services with machine learning. arXiv preprint arXiv: 2502.00865. https: //arxiv.org/pdf/2502.00865arXiv

[4] Singhal, A., Anish, P. R., Karande, S., & Ghaisas, S. (2023). Towards mitigating perceived unfairness in contracts from a non-legal stakeholder's perspective. arXiv preprint arXiv: 2312.01398. https: //arxiv.org/abs/2312.01398arXiv

[5] Schneider, P., Schopf, T., Vladika, J., Galkin, M., Simperl, E., & Matthes, F. (2022). A decade of knowledge graphs in natural language processing: A survey. Proceedings of the 2nd Conference on Automated Knowledge Base Construction. https: //aclanthology.org/2022.akbc-1.1Medium

[6] Faraoni, L. (2024). A contract law perspective on manipulative persuasive technologies: Undue influence and the digital age. University of Leeds PhD Thesis. https: //etheses.whiterose.ac.uk/id/eprint/34943/1/Faraoni_208020686_Thesis.pdf

[7] Maher, F. (2020). Algorithmic contracts and the equitable doctrine of undue influence. University of Western Australia Law Review, 47(1), 1–33. https: //research-repository.uwa.edu.au/files/92120976/SSRN_version.pdf

[8] Schwartz, A., & Ayres, I. (2021). The no-reading problem in consumer contract law. Stanford Law Review, 66(3), 545–610. https: //www.stanfordlawreview.org/wp-

DOI: 10.54254/2755-2721/2025.24509

content/uploads/sites/3/2014/03/66_Stan_L_Rev_545_AyresSchwartz.pdfstanfordlawreview.org
[9] Schneider, P., Schopf, T., Vladika, J., Galkin, M., Simperl, E., & Matthes, F. (2023). A decade of knowledge graphs in natural language processing: A survey. Medium. https: //medium.com/data-science/a-decade-of-knowledge-graphs-in-natural-language-processing-5fdb15abc2b3Medium
[10] Ayres, I., & Schwartz, A. (2021). The no-reading problem in consumer contract law. Stanford Law Review, 66(3), 545–610. https: //www.stanfordlawreview.org/wp-content/uploads/sites/3/2014/03/66_Stan_L_Rev_545_AyresSchwartz.pdf