

Risk Analysis of Engineering Procurement and Construction

Subjects: Automation & Control Systems | Computer Science, Artificial Intelligence

Contributor: Sowon CHOI, Eul-Bum Lee

The lump sum turn key (LSTK) contract for engineering, procurement, and construction (EPC) projects is a typical contract type used in large-scale and complex plant projects.

Keywords: EPC contract risk extraction ; EPC contract lexicon ; ontological semantic model ; risk level ranking

1. Introduction

The lump sum turn key (LSTK) contract for engineering, procurement, and construction (EPC) projects is a typical contract type used in large-scale and complex plant projects ^[1]. The EPC plant project combines manufacturing and services such as knowledge service, design, equipment, and construction. In addition, it is a complex industry with various front and back sectors. Furthermore, it includes global supply chains throughout the entire cycle, from bidding to maintenance ^[2]. In particular, the LSTK contract, in which the EPC contractor bears all liabilities related to design, purchase, construction, and commissioning, is an unbalanced contract as it pays more risks to the EPC contractor due to the increase in complexity when the size of the project expands ^[1]. Overseas EPC plant projects of Korean companies have been growing in earnest since the mid-2000s, and the number of orders has continued to increase due to market expansion ^[3]. However, the need to improve project risk management emerged as the EPC plant industry experienced an earnings shock, such as a decline in cut yields due to a decrease in oil prices. The development of intelligent information technology in the Fourth Industrial Revolution is currently evolving into the digital transformation of all industries. Thus, it is required to increase productivity and strengthen competitiveness using convergence technology to respond to the EPC business environment that is becoming more extensive and complex. Accordingly, the authors' research team considered applying artificial intelligence (AI) technology to manage the risk of bidding documents in the bidding stage of the project.

Invitation to bid (ITB), contract, and claim, mainly used during the EPC project, are text-based unstructured data that describe the client's requirements and significant contractual issues. Furthermore, failure to adequately review the risks of ITB in the bidding stage may result in future disputes. Nevertheless, EPC contractors struggle with ITB analysis and detection of risk clauses due to the large volume of documents, tight schedules, and lack of experienced practitioners during the bidding phase. To this end, research on a system that can analyze bidding documents, especially ITB risk factors, is required at the project bidding stage. It is necessary to convert text data in natural language form into a script that the computer recognizes. In addition, there were not many cases of NLP and AI in the EPC industry compared to other fields, such as medicine; therefore, it is relatively incomplete. Since EPC project documents consist of a large portion of unstructured text data, there is ample space for the NLP application. NLP is a branch of AI that utilizes AI to enable computers to process natural language text ^[4]. The following content explores a novel approach to automating risk analysis of EPC contracts and computational developments in NLP.

The purpose of the these content is to effectively analyze a vast amount of ITB documents in a short period and reduce the uncertainty of decision-making based on human experience and judgment. In addition, it is aimed to support the quick decision-making of EPC contractors and enhance competitiveness by automatically analyzing the critical risks of the ITB in the bidding stage of the EPC project. A novel framework of the NLP-based semantic analysis (SA) model and the bi-directional long short-term memory (bi-LSTM) method-based risk level ranking (RLR) model is proposed to analyze the contract risk clause of EPC ITB automatically.

The proposed SA model is an approach that applies the EPC contract lexicon to SVO tuples to develop semantic rules. Then, it extracts risks according to whether the analysis target sentence matches the rules. The following content applied the ontology-based semantic information extraction (IE) technique, which maps heterogeneous contract clauses to ontology-based lexicons. Ontology expresses the relationship between objects in a form that a computer can process, and by linking domain knowledge, it becomes the basis for developing semantic rules. In ontology-based semantic IE, the

lexicon configuration is significant because it determines the risk clauses by considering the semantic relationship of sentence elements based on the information stored in the lexicon, rather than using a simple keyword search method. Therefore, ontology-based semantic IE performs better than syntactic IE [5]. In addition, a PDF structuralization module that recognizes and formalizes text data in documents separately was designed to improve the accuracy of text data analysis. The RLR model was created to address the issue of using the bi-LSTM algorithm, check the risk of each sentence in the ITB, and classify the risk class. Furthermore, the RLR model classifies and extracts each sentence of the EPC contract document into five levels according to the degree of risk. Moreover, a dataset for model training was developed, and hyperparameters were optimized to maximize model performance.

2. Knowledge-Based Risk Extraction for EPC Projects

Ebrahimnejada et al. [6] proposed the extended VIKOR method based on the fuzzy set theory as a new risk evaluation approach in large-scale projects. They applied it to the Iranian power plant project to compare the differences with the traditional version. Hung and Wang [7] conducted a study to identify the main risk factors that cause delays in hydropower construction projects in Vietnam and analyze the degree of impact of each risk factor on construction. Jahantigh and Malmir [8] identified, evaluated, and prioritized significant financial risks of EPC projects in terms of national development in developing countries. Furthermore, their work was based on the fuzzy TOPSIS model and they applied the refinery project as a case study. Kim et al. [9] developed the Detail Engineering Completion Rating Index System (DECRIS) that minimizes the rework of EPC contractors and supports schedule optimization for offshore EPC projects. This model improved existing theories, such as the Project Definition Rating Index (PDRI) and front end loading (FEL). Their study verified the effect of schedule and cost through 13 megaprojects. Kabirifar and Mojtahedi [10] studied the most critical factors in EPC project execution by applying the TOPSIS method to a large-scale residential construction project in Iran. In addition, they derived that procurement is the most vital risk factor. Gunduz and Almuajebh [11] ranked 40 critical success factors (CSFs) after reviewing the literature on CSFs considering stakeholder impacts in construction projects. Their collected data were analyzed using the relative importance index (RII) and analytic hierarchy process (AHP) method with Saaty random index. Koulinas et al. [12] proposed a simulation-based approach to estimate the project schedule's delay risk and predict in-time project completion. This approach, implemented through a hotel renovation project, showed better uncertainty expression and superior predictions in comparison to the classic PERT method when estimating budget and time-critical overruns. Okudan et al. [13] developed a knowledge-based risk management tool (namely, CBLisk) using case-based reasoning (CBR). As a web-based tool, this system is characterized by applying the project similarity list in the form of fuzzy linguistic variables for effective case search.

3. Automatic Extraction of Contract Risks Using AI Technology in EPC Projects

In recent years, research on extracting contract risk from legal documents has been actively conducted by applying AI technology. Surden [14] studied the method of representing specific contractual obligations in computer data for financial contracts, such as stock option contracts. Automated manual comparison has significantly reduced transaction costs associated with contract monitoring compared to traditional written contracts as it applies a technology that transforms specific contract terms into a set of computer-processable rules. In 2018, LawGeex [15] collaborated with 20 experienced lawyers educated in the United States to conduct a study of a contract review platform developed with the AI application. The study, which looked at non-disclosure agreements (NDAs), showed that AI was 94 percent accurate compared to experienced lawyers, who were 85 percent accurate. Their study improved the quality of legal human resources through faster and more reliable contract management. Cummins and Clack [16] reviewed the concept of “computable contracts”, which both humans and computers can understand as the concept exists in text form in natural language. Furthermore, they proposed an integrated framework of various technologies and approaches to model their concepts. Dixon Jr. [17] described the application cases of various AI technologies used in the legal field, such as crime prediction, prevention, detection, and contract drafting and review. Clack [18] studied the problems of converting natural language into computer code that occurred when developing a “smart legal contract”, which automates legal contracts using computer technology. His study explained the importance of language design in smart contracts, such as computable language, natural language, and the meaning of the language expression. Salama and El-Gohary [19] studied an automated compliance-checking model that applied deontic logic to the construction domain.

EPC documents consist of a significant portion of text-based unstructured data, while NLP technology is mainly used for text information extraction and retrieval. NLP is an AI-related field of human–computer interaction that enables a computer to interpret human language through machine learning [20]. Zhang and El-Gohary [21] presented a semantic rule-based NLP approach using information extraction (IE) from complex construction regulations. Their study was meaningful as it

allowed an advancement in the existing method of selectively extracting only some information from documents. Williams and Gong [22] proposed a risk model to predict cost overruns using data-mining and classification algorithms in bidding documents for construction projects. However, there was a limitation in analyzing only simple keyword-oriented text data, such as project summary information for text analysis. Lee and Yi [23] proposed a bidding risk prediction model using construction project bidding information text mining. However, there was no quantitative explanation of how much the cost should reflect. Zoua et al. [24] proposed an approach that combines two NLP techniques, a vector space model (VSM) and semantic query expansion, to improve search efficiency for accident cases in a construction project. As a result of the study, the problem of semantic similarity remains a significant challenge. Lee et al. [25] proposed a contract risk extraction model for construction projects by applying NLP's automatic text analysis method to the Fédération Internationale Des Ingénieurs-Conseils (FIDIC) Redbook. Their study showed the performance of extracting only about 1.2 percent of the whole sentence as a risk, and their model cannot be applied to other types of contracts other than FIDIC-based, such as offshore plants. Moon et al. [26] proposed an information extraction framework that used Word2Vec and named entity recognition (NER) to develop an automatic review model for construction specifications when bidding for infrastructure projects. Their model targeted only the text data of the construction specification document and it could not analyze the text data shown in the tables or drawings included in the document. Choi et al. [27] developed the Engineering Machine Learning Automation Platform (EMAP). This integrated platform supports decision-making by applying AI and machine learning (ML) algorithms based on data generated throughout the EPC project cycle. Their study is meaningful because it is the first integrated platform for risk extraction of the entire EPC project life cycle. Choi et al. [28] developed a model for checking the presence of a risk clause in an EPC contract using NER and a phrase-matcher. Park et al. studied an ML-based model to extract technical risks from EPC technical specification documents [29]. Choi et al. [27] and Park et al. [29] were interrelated as they created the parts of the sub-element constituting the *EMAP* system. Fantoni et al. [30] utilized state-of-the-art computer language tools with an extensive knowledge base to automatically detect, extract, split, and assign information from technical documents when tendering for a railway project. The implementation of the methodology was utilized during a high-speed train project.

4. Text Classification

Text classification classifies text data into meaningful categorical classes and is one of the leading research areas of NLP [31]. Traditional text classification methods include dictionary-based and basic machine learning methods [31]. Since the 2000s, it has been replaced by deep learning such as recurrent neural network (RNN), long short-term memory (LSTM), and convolutional neural network (CNN) [32]. Currently, a more powerful text classification technique, such as BERT, has emerged [33]. RNN is one of the neural network architectures used for text mining and classification. Additionally, RNN is a kind of artificial neural network in which directed edges connect hidden nodes to form a directed cycle [34]. Furthermore, it is suitable for processing time-series data that appear sequentially, such as speech and text [35]. However, RNNs have a problem of long-term dependencies in which past learning results disappear. Thus, LSTM was designed to overcome this issue of RNNs [36][37]. The LSTM model proposed by Hochreiter and Schmidhuber [37] is internally controlled by the gating mechanism called input gate, output gate, and forget gate. By improving the long-term dependency problem of RNN, it processes massive data such as time-series data without any problem. However, the unidirectional LSTM has the disadvantage of preserving only past information [38]. Schuster and Paliwal [39] proposed a bi-LSTM model that extends the unidirectional LSTM through introducing a second hidden layer to compensate for this problem of LSTM. Bi-LSTM uses LSTM cells in both directions, therefore past and future information can be exploited [39]. In addition, it is mainly used for text classification due to its excellent performance on sequential modeling problems [33]. Li et al. [40] reviewed text classification methods from 1961 to 2021 and created a taxonomy for text classification tasks from traditional models to deep learning. They also introduced the datasets with a summary table and provided the quantitative results of the leading models. Minaee et al. [41] provided a comprehensive review of deep-learning-based models for text classification developed in recent years and discussed their technical contributions, similarities, and strengths. They also explained a summary of more than 40 popular datasets for text classification.

References

1. DLA Piper. EPC Contracts in the Process Plant Sector. Available online: www.dlapiper.com (accessed on 5 February 2022).
2. Ritsche, F.-P.; Wagner, R.; Schlemmer, P.; Steinkamp, M.; Valnion, B.D. Innovation Project EPC 4.0 'Unleashing the Hidden Potential'; ProjectTeam: Hamburg, Germany, 2019.
3. International Trade Administration. South Korea-Construction Services. Available online: <https://www.trade.gov/country-commercial-guides/south-korea-construction-services> (accessed on 7 February 2022).

4. Vogl, R. *The Coming of Age of Legal Technology*; Stanford University: Stanford, CA, USA, 2016.
5. Lane, H.; Hapke, H.; Howard, C. *Natural Language Processing in Action: Understanding, Analyzing, and Generating Text with Python*; Simon and Schuster: New York, NY, USA, 2019.
6. Ebrahimnejad, S.; Mousavi, S.M.; Tavakkoli-Moghaddam, R.; Heydar, M. Evaluating high risks in large-scale projects using an extended VIKOR method under a fuzzy environment. *Int. J. Ind. Eng. Comput.* **2012**, *3*, 463–476.
7. Hung, M.S.; Wang, J. Research on Delay Risks of EPC Hydropower Construction Projects in Vietnam. *Int. J. Power Energy Eng.* **2016**, *4*, 8.
8. Jahantigh, F.F.; Malmir, B.; Avilaq, B.A. Economic risk assessment of EPC projects using fuzzy TOPSIS approach. *Int. J. Ind. Syst. Eng.* **2017**, *27*, 161–179.
9. Kim, M.-H.; Lee, E.-B.; Choi, H.-S. Detail Engineering Completion Rating Index System (DECRIS) for Optimal Initiation of Construction Works to Improve Contractors' Schedule-Cost Performance for Offshore Oil and Gas EPC Projects. *Sustainability* **2018**, *10*, 2469.
10. Kabirifar, K.; Mojtahedi, M. The impact of Engineering, Procurement and Construction (EPC) Phases on Project Performance: A Case of Large-scale Residential Construction Project. *Buildings* **2019**, *9*, 15.
11. Gunduz, M.; Almuajebh, M. Critical Success Factors for Sustainable Construction Project Management. *Sustainability* **2020**, *12*, 1990.
12. Koulinas, G.K.; Xanthopoulos, A.S.; Tsilipiras, T.T.; Koulouriotis, D.E. Schedule delay risk analysis in construction projects with a simulation-based expert system. *Buildings* **2020**, *10*, 134.
13. Okudan, O.; Budayan, C.; Dikmen, I. A knowledge-based risk management tool for construction projects using case-based reasoning. *Expert. Syst. Appl.* **2021**, *173*, 114776.
14. Surden, H. Computable contracts. *UC Davis Law Rev.* **2012**, *46*, 629.
15. LawGeex. Comparing the Performance of AI to Human Lawyers in the Review of Standard Business Contracts. Available online: <https://ai.lawgeex.com/rs/345-WGV-842/images/LawGeex%20eBook%20AI%20vs%20Lawyers%202018.pdf> (accessed on 10 January 2022).
16. Cummins, J.; Clack, C. Transforming Commercial Contracts through Computable Contracting. *arXiv* **2020**, arXiv:2003.10400.
17. Dixon, H.B., Jr. What judges and lawyers should understand about artificial intelligence technology. *ABA J.* **2020**, *59*, 36–38.
18. Clack, C.D. Languages for Smart and Computable Contracts. *arXiv* **2021**, arXiv:2104.03764.
19. Salama, D.A.; El-Gohary, N.M. Automated compliance checking of construction operation plans using a deontology for the construction domain. *J. Comput. Civ. Eng.* **2013**, *27*, 681–698.
20. Chopra, D.; Joshi, N.; Mathur, I. *Mastering Natural Language Processing with Python*; Packt Publishing Ltd.: Birmingham, UK, 2016.
21. Zhang, J.; El-Gohary, N. Automated reasoning for regulatory compliance checking in the construction domain. In *Construction Research Congress 2014: Construction in a Global Network*; ASCE: Reston, VA, USA, 2014; pp. 907–916.
22. Williams, T.P.; Gong, J. Predicting construction cost overruns using text mining, numerical data and ensemble classifiers. *Autom. Constr.* **2014**, *43*, 23–29.
23. Lee, J.; Yi, J.-S. Predicting project's uncertainty risk in the bidding process by integrating unstructured text data and structured numerical data using text mining. *Appl. Sci.* **2017**, *7*, 1141.
24. Zou, Y.; Kiviniemi, A.; Jones, S.W. Retrieving similar cases for construction project risk management using Natural Language Processing techniques. *Autom. Constr.* **2017**, *80*, 66–76.
25. Lee, J.H.; Yi, J.-S.; Son, J.W. Development of automatic-extraction model of poisonous clauses in international construction contracts using rule-based NLP. *J. Comput. Civ. Eng.* **2019**, *33*, 04019003.
26. Moon, S.; Lee, G.; Chi, S.; Oh, H. Automated construction specification review with named entity recognition using natural language processing. *J. Constr. Eng. Manag.* **2021**, *147*, 04020147.
27. Choi, S.-W.; Lee, E.-B.; Kim, J.-H. The Engineering Machine-Learning Automation Platform (EMAP): A Big-Data-Driven AI Tool for Contractors' Sustainable Management Solutions for Plant Projects. *Sustainability* **2021**, *13*, 10384.
28. Choi, S.J.; Choi, S.W.; Kim, J.H.; Lee, E.-B. AI and Text-Mining Applications for Analyzing Contractor's Risk in Invitation to Bid (ITB) and Contracts for Engineering Procurement and Construction (EPC) Projects. *Energies* **2021**, *14*, 4632.

29. Park, M.-J.; Lee, E.-B.; Lee, S.-Y.; Kim, J.-H. A Digitalized Design Risk Analysis Tool with Machine-Learning Algorithm for EPC Contractor's Technical Specifications Assessment on Bidding. *Energies* 2021, 14, 5901.
30. Fantoni, G.; Coli, E.; Chiarello, F.; Apreda, R.; Dell'Orletta, F.; Pratelli, G. Text mining tool for translating terms of contract into technical specifications: Development and application in the railway sector. *Comput. Ind.* 2021, 124, 103357.
31. Jang, B.; Kim, M.; Harerimana, G.; Kang, S.-U.; Kim, J.W. Bi-LSTM model to increase accuracy in text classification: Combining Word2vec CNN and attention mechanism. *Appl. Sci.* 2020, 10, 5841.
32. Pouyanfar, S.; Sadiq, S.; Yan, Y.; Tian, H.; Tao, Y.; Reyes, M.P.; Shyu, M.-L.; Chen, S.-C.; Iyengar, S. A survey on deep learning: Algorithms, techniques, and applications. *ACM Comput. Surv. (CSUR)* 2018, 51, 1–36.
33. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* 2019, arXiv:1810.04805.
34. Zhou, S.K.; Rueckert, D.; Fichtinger, G. *Handbook of Medical Image Computing and Computer Assisted Intervention*; Academic Press: Cambridge, MA, USA, 2019.
35. Kowsari, K.; Jafari Meimandi, K.; Heidarysafa, M.; Mendu, S.; Barnes, L.; Brown, D. Text Classification Algorithms: A Survey. *Information* 2019, 10, 150.
36. Pascanu, R.; Mikolov, T.; Bengio, Y. On the difficulty of training recurrent neural networks. In *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, GA, USA, 17–21 June 2013; pp. 1310–1318.
37. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural. Comput.* 1997, 9, 1735–1780.
38. Schuster, M.; Paliwal, K.K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* 1997, 45, 2673–2681.
39. Zhou, P.; Qi, Z.; Zheng, S.; Xu, J.; Bao, H.; Xu, B. Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling. *arXiv* 2016, arXiv:1611.06639.
40. Li, Q.; Peng, H.; Li, J.; Xia, C.; Yang, R.; Sun, L.; Yu, P.S.; He, L. A Survey on Text Classification: From Traditional to Deep Learning. *ACM Trans. Intell. Syst. Technol.* 2022, 13, 364.
41. Minaee, S.; Kalchbrenner, N.; Cambria, E.; Nikzad, N.; Chenaghlu, M.; Gao, J. Deep Learning—Based Text Classification. *ACM Comput. Surv.* 2022, 54, 1–40.

Retrieved from <https://encyclopedia.pub/entry/history/show/58463>