Discovery of potential technological opportunities in the green building field from a technology convergence perspective: a two-stage prediction approach based on interpretable machine learning

Shuai Feng Kaijian Li Guiwen Liu *

[1] PhD Student, School of Management Science and Real Estate, Chongqing University, China

[2] Faculty, School of Management Science and Real Estate, Chongqing University, China

[3] Vice-President, Chongqing University, China

Corresponding author's E-mail: gwliu@cqu.edu.cn

Abstract

As one of the world's largest energy consumers and carbon emitters, the construction industry plays a pivotal role in advancing global sustainability. Promoting technological innovation is essential for the industry's green transformation while maintaining development goals. However, current research reveals a significant gap in identifying and assessing potential technological innovation opportunities within the green construction sector, resulting in a lack of decision-making guidance for governments and innovators during the research and development phase. Recognizing this, the study proposes a two-stage technology opportunity prediction approach based on interpretable machine learning from the perspective of technology convergence. Unlike traditional methods, this approach estimates the likelihood of emerging technological opportunities and anticipates the impact of convergence events. By examining 600,442 patent documents related to green technologies and construction, the study forecasts potential technology convergence innovations and investigates the factors driving these trends. The findings provide critical decision support for policymakers and organizations to develop strategies for green technology innovation.

Keywords: Technology Convergence; Technology Prediction; Interpretable Machine Learning; Green Technology Innovation; Construction Industry

1. INTRODUCTION

Identifying opportunities for green technologies (GTs) in the construction industry presents significantly more complexities than recognizing innovations in traditional construction technologies. While various approaches exist, they exhibit notable limitations. First, these methods predominantly target high-tech industries such as bioinformatics (Kwon & Sohn, 2022), advanced materials (Sasaki & Sakata, 2021), autonomous vehicles (Kim & Sohn, 2020), and smart health (Wang & Lee, 2023), with limited application in traditional, resource-driven sectors like construction. The efficacy of current predictive models remains unvalidated within the construction context. Second, existing technology convergence discovery methods often neglect the distinct characteristics and development patterns across different industries. Empirical research shows substantial variations in the technological development trajectories of various sectors (Corrocher et al., 2007; Krafft et al., 2011). The construction industry, characterized by the integration of diverse knowledge bases and multi-agent collaborations, requires consideration of unique indicators to capture its innovation dynamics (Herrera Rodrigo et al., 2020). Third, many current studies overlook the need for interpretability in predictive models, resulting in opaque decision-making processes and limiting the models' applicability and generalizability. Lastly, previous methods primarily emphasize predicting the likelihood of technology convergence, without assessing the potential impact of these convergence points. This represents a critical gap, as not all convergence points yield significant innovations (Pezzoni et al., 2022; Ardito et al., 2023)

To address these limitations, this study proposes a two-stage machine learning-based approach to predict convergence opportunities: (1) estimating the probability of technology convergence and (2) forecasting its impact. To enhance predictive accuracy, a novel set of indicators tailored to the technological features of the construction industry was developed. The SHAP method was employed to provide interpretability, elucidating the model's internal mechanics. Lastly, the study identifies potential green technology innovations (GTIs) within the construction sector.

2. LITERATURE REVIEW

GTs in construction represent significant technological advancements aimed at environmental protection and enhancing sustainability, drawing widespread global attention (Ahmad et al., 2019). These technologies are designed to reduce energy consumption, promote health, and support sustainable development goals across all phases of construction projects (Chan Albert et al., 2017). Beyond traditional construction practices, GTs have emerged as strategic components essential for resource conservation (Yu et al., 2023), energy reduction (Du & Li, 2019), waste minimization (Yin & Li, 2019), and environmental quality improvement (Devine & McCollum, 2019; Hwang et al., 2017). They are now indispensable in accelerating societal sustainability and constitute a significant focus of research (Lai et al., 2017; Chan Albert et al., 2017; Lai et al., 2023).

The existing literature predominantly addresses the benefits of GTs in construction and the factors influencing their adoption. For instance, Besir and Cuce (2018) reviewed the characteristics and advantages of various GTs, while Yang et al. (2021) used an evolutionary game optimization model to explore changes in stakeholder strategies. Chan et al. (2017) identified key barriers to GT adoption, such as resistance to change and high costs. Similarly, Yin and Li (2019) highlighted research and development expenses as significant obstacles, and Chan et al. (2018) noted that the high costs associated with GTs impede their adoption in both developed and developing regions.

Among the factors limiting GT development, economic costs stand out as the most significant constraint (Zhang et al., 2021). Efficiently identifying high-potential technology convergence opportunities, minimizing ineffective research and development investments, and optimizing the allocation of innovation resources are crucial steps toward advancing GTs (Zhang et al., 2021). However, current approaches to discovering GT opportunities in construction remain limited, often relying on expert assessments (Besir & Cuce, 2018; Lai et al., 2023) rather than quantitative, data-driven methods. This reliance results in a lack of comprehensive information for decision-makers engaged in technological innovation.

3. METHODOLOGY

3.1. Data Collection

This study gathered patent data from the IncoPat database, focusing on patents granted in the green and construction fields in China from 2013 to 2021. To minimize retrieval omissions caused by semantic differences, the search employed an International Patent Classification (IPC) code retrieval strategy, with green patent criteria based on IPC codes for GTs provided by the WIPO GREEN Inventory database.

3.2. Indicator Construction

Building on previous studies of technology convergence prediction, three dimensions—knowledge network, bibliometrics, and semantics—were identified for constructing the predictive indicators. First, the generation of new technologies is shaped by the knowledge network, where existing knowledge elements combine or recombine (Smojver et al., 2021). Second, bibliometric data, which records the technological development context, provides valuable insights into the precursors of technology convergence (Karvonen & Kässi, 2013). Lastly, semantic information from patents serves as supplementary data, with higher semantic similarity scores indicating potential technology convergence (Kim & Sohn, 2020). Recognizing that the advancement of GTs in construction is frequently shaped by collaborative interactions among innovation subjects, we gathered data from the collaboration network. This data was then integrated with the knowledge network, creating a "knowledge-collaboration" dual-layer network to provide additional insights.

3.3. Model Construction and Testing

2.3.1 Technology Convergence Probability Prediction Model

This study treated technology convergence probability prediction as a binary classification problem, aiming to determine whether two knowledge elements would converge. A set of 22 indicators was curated, and five commonly used machine learning algorithms in management studies—Support Vector Machine (SVM), Artificial Neural Network (ANN), Decision Tree (DT), Random Forest (RF), and Gradient Boosting Decision Tree (GBDT)—were employed to build the model (Choudhury et al., 2021). Model evaluation was conducted using six performance metrics to identify the optimal algorithm for predicting technology convergence probability.

2.3.2 Technology Convergence Impact Prediction Model

Following the identification of potential technology convergence opportunities, the subsequent step involved forecasting the impact of these opportunities, recognizing that not all convergence points result in significant technological advancements. Technical impact reflects the extent to which a technology prospers (Ardito et al., 2023) and is indicated by the number of subsequent invention patents arising from a given convergence opportunity (Pezzoni et al., 2022). Thus, predicting the impact was approached as a regression task, aiming to forecast the number of patents generated based on the convergence opportunity. In addition to the original 22 indicators, the metric "Historical Impact" was introduced, based on the technological lifecycle theory, which posits that impact levels follow a lifecycle curve (Hoppmann et al., 2020; Pezzoni et al., 2022). The regression models were constructed using the same five algorithms: SVM, ANN, DT, RF, and GBDT (Choudhury et al., 2021).

3.4. Interpretability Analysis

The SHAP (Shapley Additive exPlanations) method, grounded in cooperative game theory (Shapley, 1953), is recognized for its rigorous theoretical basis. Developed by Lundberg and Lee (2017), SHAP quantifies the contribution of each feature to a prediction by comparing outcomes with and without the feature, effectively isolating its impact (Lundberg et al., 2020). This approach enhances model transparency and aids in validating the trustworthiness and fairness of the predictions.

4. RESULTS AND DISCUSSION

4.1. Model Performance

The research results indicate that the models perform exceptionally well in discovering technology convergence opportunities. **Table 1** compares the performance of machine learning models for technology convergence probability prediction using Accuracy, Precision, Recall, F1, AUC, and Precision@k as evaluation metrics.

Table 1 Performance of technology convergence probability prediction model.

Metrics	SVM	ANN	DT	RF	GBDT
Accuracy	0.84	0.85	0.82	0.84	0.84
Precision	0.77	0.81	0.79	0.83	0.82
Recall	0.71	0.68	0.61	0.63	0.63
F1	0.74	0.74	0.69	0.72	0.71
AUC	0.90	0.91	0.88	0.91	0.91
Precision@k	0.91	0.94	0.91	0.93	0.93

Table 2 provides a comparative analysis of various machine learning models using MAE, Root RMSE, and R^2 as evaluation metrics.

Table 2 Performance of technology convergence probability prediction model.

Metrics	SVM	ANN	DT	RF	GBDT
MAE	2.12	2.04	3.43	2.34	2.12
RMSE	117.86	69.89	207.06	102.80	117.86
R^2	0.69	0.82	0.45	0.73	0.76

4.2. Model Interpretation

3.2.2 Interpretability Analysis of the Probability Prediction Model

The top ten indicators influencing the likelihood of technology convergence include *Katz, Salton, Semantics_ti, HPI, PA, ACT, Proximity, LP, Scale, and HDI*. These findings align with previous studies, which emphasize the importance of similarity measures such as *CN, HDI, HPI, LHN, Salton, PA, Katz, ACT, MFI, LP, Semantics_ti, and Semantics_ab* in forecasting technology convergence (Kim & Sohn, 2020). An increase in network structure complexity and semantic similarity between knowledge elements is considered a precursor to technology convergence (Kim & Sohn, 2020).

3.2.2 Interpretability Analysis of the Impact Prediction Model

The most significant indicators for predicting the impact of technology convergence opportunities include *Historical Impact, Proximity, Katz, MFI, ACT, Scale, Salton, Semantics_ti, Semantics_ab, and Sorenson*, with *Historical Impact* standing out as particularly influential. It demonstrates a strong positive correlation with the impact of convergence opportunities, being more than four times as influential as the next most important indicator, Proximity. This pattern aligns with the principles of the technology lifecycle, which reflect the developmental trends of technological impact (Gal et al., 2022). In contrast, indicators such as *Scale and Complexity*, along with most *similarity measures* (e.g., *Katz, MFI, ACT, Salton, Semantics_ti, Semantics_ab, Sorenson, PA, and AA*), generally show a negative correlation with the impact of technology convergence opportunities.

4.3. Technology Convergence Opportunities Predicted in the Future

Several convergence pairs, including (E03B, A01G), (E02B, B63B), (E03F, B01D), (E01C, C04B), (E02D, A01G), (E03F, A01G), (E03B, B01D), and (E02B, A01G), exhibit not only high probabilities of occurrence but also significant levels of impact. These 'star' convergence opportunities present higher returns on investment and lower development risks, offering valuable creative inspiration for researchers.

5. CONCLUSIONS

Green technology innovation is essential for achieving sustainable development in the construction industry while sustaining economic growth. Technology convergence facilitates the green transformation of construction technology by integrating green knowledge with traditional building technologies through technological redevelopment. Given the high uncertainty and complexity involved in selecting technological paths, a data-driven approach to discovering technology convergence opportunities is necessary to support innovation decision-making (Zhang et al., 2021).

This study offers a dual contribution. Academically, it advances previous research by proposing a two-stage technology convergence opportunity prediction approach, which not only estimates the probability of technology convergence but also evaluates the potential impact of such opportunities. The study introduces tailored innovation indicators specific to the technological development characteristics of the construction industry and assesses their effectiveness. By employing the SHAP method, the study elucidates the model's decision-making process, highlighting the roles and interactions of the indicators, and thereby addressing the interpretability limitations of traditional machine learning models. Practically, the study predicts potential technology convergence opportunities between the construction and green sectors for 2022 – 2024. It identifies key technology convergence opportunities, aiming to reduce the time and costs associated with expert assessments, allowing experts to focus more on the detailed technical evaluation of high-impact ('star') opportunities.

Despite these contributions, the study has limitations. First, the reliance on patent databases for discovering technology opportunities may not fully capture trends in basic research and market developments. Future research could integrate diverse data sources, such as technology news, academic publications, and user communities, to enhance the analysis. Additionally, the study focuses on convergence between two knowledge domains, while many technological innovations involve the integration of multiple knowledge elements. Future work could explore predicting technology convergence patterns across multiple domains. Lastly, developing expert-machine collaborative decision-making models could help mitigate the data biases inherent in using machine learning algorithms alone.

6. REFERENCES

Ahmad, Tayyab, Aibinu, Ajibade Ayodeji, & Stephan, André. (2019). Managing Green Building Development – a Review of Current State of Research and Future Directions. Building and Environment, 155, 83-104. https://doi.org/10.1016/j.buildenv.2019.03.034

Kwon, D., & Sohn, S. Y. (2024). Convergence Technology Opportunity Discovery for Firms Based on Technology Portfolio Using the Stacked Denoising Autoencoder (Sdae). IEEE Transactions on Engineering Management, 71, 1804-1818. https://doi.org/10.1109/TEM.2022.3208871

Summa, S., Remia, G., Di Perna, C., & Stazi, F. (2023). Experimental and Numerical Study on a New Thermal Masonry Block by Comparison with Traditional Walls. Energy and Buildings, 292, 113125. https://doi.org/10.1016/j.enbuild.2023.113125

Zhang, Jingxiao, Ouyang, You, Ballesteros-Pérez, Pablo, Li, Hui, Philbin, Simon P., Li, Zhaoling, & Skitmore, Martin. (2021). Understanding the Impact of Environmental Regulations on Green Technology Innovation Efficiency in the Construction Industry. Sustainable Cities and Society, 65, 102647. https://doi.org/10.1016/j.scs.2020.102647