

Machine Learning Applications in Facility Life-Cycle Cost Analysis: A Review

Xinghua Gao¹; Pardis Pishdad-Bozorgi, Ph.D.²; Dennis R. Shelden, Ph.D.³; and Yuqing Hu⁴

¹School of Building Construction, Georgia Institute of Technology, 280 Ferst Dr., Atlanta, GA 30332. E-mail: gaoxh@gatech.edu

²School of Building Construction, Georgia Institute of Technology, 280 Ferst Dr., Atlanta, GA 30332. E-mail: pardis.pishdad@design.gatech.edu

³Digital Building Laboratory, School of Architecture, Georgia Institute of Technology, 245 4th St. NW, Atlanta, GA 30332. E-mail: dennis.shelden@design.gatech.edu

⁴School of Building Construction, Georgia Institute of Technology, 280 Ferst Dr., Atlanta, GA 30332. E-mail: yuqing_hu@gatech.edu

ABSTRACT

A large amount of resources are spent on constructing new facilities and maintaining the existing ones. The total cost of facility ownership can be minimized by focusing on reducing the facilities life-cycle costs (LCCs) rather than the initial design and construction costs. With the developments of machine learning in predictive analytics and the utilizing building systems that provide ubiquitous sensing and metering devices, new opportunities have emerged for architecture, engineering, construction, and operation (AECO) professionals to obtain a deeper level of knowledge on buildings' LCCs. This paper provides a state-of-the-art overview of the various machine learning applications in the facility LCC analysis field. This paper aims to present current machine learning for LCC research developments, analyze research trends, and identify promising future research directions.

Keywords: Life Cycle Cost Analysis (LCCA), Machine Learning, Facilities Management, Cost Prediction, Data Mining

INTRODUCTION

Because of the long life spans of buildings, robust decisions regarding the economic efficiency of alternative materials, components, and systems demand a full lifecycle perspective that goes beyond the initial cost and regular maintenance and repair (Noshadravan et al., 2017). Hence, the Life Cycle Cost Analysis (LCCA) has become increasingly important in new building design and existing building retrofitting, refurbishment, and renovations. In recent years, the developments of machine learning techniques provide building experts with new opportunities to achieve more accurate predictions of facility-related costs in the early design phase or even programming phase. To our knowledge, there is no comprehensive review that specifically focuses on recent research of machine learning for facility life-cycle cost (LCC) prediction. We try to close this gap with the present paper. Through a literature review, this paper aims to synthesize and presents a summary of current machine learning techniques for LCC research developments, analyze research trends, and identify promising future research directions.

Building-related costs usually fall into the following categories (Fuller, 2010): 1) initial costs – purchase, acquisition, design and construction costs, 2) utility costs – electricity, water, gas, and garbage costs, 3) operation, maintenance, and repair (O&M) costs, 4) replacement costs – capital replacements of building systems that have different service lives, 5) residual values – resale or salvage values or disposal costs, 6) finance charges – loan interest payments, and 7) non-monetary benefits or costs – such as the benefit derived from a quiet HVAC system or

improved lighting. In this paper, we investigate the publications related to the first three categories, which are the major components of a building's LCCs.

REVIEW APPROACH

To identify related publications involving machine learning applications in the facility LCC prediction field, a keyword search is performed in academic databases, including ELSEVIER, EMERALD, EBSCO, WILEY, ASCE, CIB, SPRINGER, T&F, and ISPRS. Articles with abstracts contain “machine learning” or “prediction” and the keywords “building cost”, “energy consumption”, “operation cost”, and “maintenance cost” are identified and reviewed. 88 publications related to our research topic are identified and reviewed. Each reviewed paper is examined in the following aspects: 1) research methodology, 2) algorithm used, 3) applicable facility type, 4) what kind(s) of costs are considered, 5) what descriptive attributes are used in the prediction model, 6) has a case study/experiment or not, 7) the size of data set, 8) challenges and difficulties discussed, and 9) suggested future research. The results and findings from the analysis, and the recommendations for future research are provided in the following sections.

OVERVIEW OF MACHINE LEARNING APPLICATIONS IN BUILDING COST PREDICTION

Initial Construction Costs Prediction

Accurate estimation in the early design stage is vital for the successful execution of a construction project. Using machine learning techniques, research studies have provided practitioners with decision-support tools for estimating construction duration and costs before the completion of a project's design stage, or even during the programming phase (C. W. Koo et al., 2010; Hong et al., 2011; Jin et al., 2016). The construction costs prediction studies can be categorized into three major groups based on the methods used, which are 1) regression analysis (Trost & Oberlender, 2003; H. Li et al., 2005; Zayed & Halpin, 2005; Lowe et al., 2006; Sonmez, 2008; Jafarzadeh et al., 2014; Alshamrani, 2017) 2) Case-based Reasoning (Dogan et al., 2006, 2008; C. W. Koo et al., 2010; Hong et al., 2011; Jin et al., 2016), and 3) Artificial Neural Network (Kim, Yoon, et al., 2004; Shi & Li, 2008; Cheng et al., 2010; Bala et al., 2014; Dursun & Stoy, 2016).

Studies have been conducted to compare the cost prediction performance of models based on different machine learning methods. For example, Kim et al. (2004) compared the accuracy of Multiple Regression Analysis (MRA), Artificial Neural Network (ANN), and Case-based Reasoning (CBR) by experimenting on 530 residential buildings' construction costs. The results indicated that although the ANN model yields more accurate results than the MRA and the CBR models, the CBR model performed better than the ANN model in terms of ease of updating and consistency in the variables stored for long-term use. Researchers have also studied the performance of machine learning methods in specific cost prediction cases. Based on 71 projects conducted by a medium sized electrical contractor, Aibinu et al. (2015) concluded that the cost forecasting models based on ANN outperform regression models in predicting the costs of light wiring, power wiring, and cable pathways. Sajadfar & Ma (2015) compared the prediction accuracies of the models based on Linear Regression, Multilinear Regression, K-Nearest Neighbors (KNN) Regression, Decision Tree Regression, and ANN. They found that the ANN model shows to be the most accurate for welding operations.

The most commonly used descriptive attributes for developing the construction cost

prediction models in above-mentioned references involve 1) building floor area, 2) number of floors, 3) structure type, 4) number of rooms, 5) roof type, 6) foundation type, 7) topography & soil condition, and 8) construction duration.

Utility Consumption Prediction

Understanding the underlying dynamics of building utility consumption (energy, water, and gas) and predicting the consumption are essential for building resource planning, management, and conservation (Amasyali & El-Gohary, 2018; Zhang et al., 2018). Energy (electricity) consumption prediction is the most extensively studied topic in the facility LCC prediction field. This is probably because the electricity meters and sensors distributed in facilities provide sufficient high-resolution data – hourly or even quarter-hourly – for researchers to investigate the utility costs in detail (Moon et al., 2018; Park et al., 2018; Sala-Cardoso et al., 2018). The most commonly used machine learning methods for energy forecasting involve: 1) Artificial Neural Network (Mocanu et al., 2016; Park et al., 2018; Sala-Cardoso et al., 2018), 2) Support Vector Machines (SVM) Regression (Jain et al., 2014; Chou & Ngo, 2016), and 3) Case-based Reasoning (An et al., 2007; Ji et al., 2014).

Most of the reviewed studies in the utility consumption prediction field developed multiple machine learning models and compared their performance (Robinson et al., 2017; Bouktif et al., 2018). For example, Geysen et al. (2018) developed a thermal load forecasting system that incorporates a collection of machine learning methods – linear regression (LR), extremely randomized trees regression (ETR), ANN, and SVM regression. The experiment results indicated that the LR performs worst while the ANN and ETR are slightly better than the SVM. The study conducted by Moon et al. (2018) also showed that the ANN-based model outperforms the SVM regression-based model in electric load forecasting. However, Idowu et al. (2016)'s study showed SVM gave the best prediction performance compared to ANN and multiple linear regression in forecasting the thermal load in district heating substations.

Although attribute importance (weight) depends on the specific machine learning model, yet certain attributes will always dominate the attribute space (Zhang et al., 2018). The most commonly used descriptive attributes for utility consumption models are 1) building age, 2) building function/type, 3) building floor area, and 4) number of floors. Advances in machine learning techniques enabled researchers to develop prediction models without a large quantity of data. Li et al. (2017) proposed an extreme deep learning approach that can extract most influential building energy consumption attributes and improve the prediction accuracy.

Operation and Maintenance Costs Prediction

Studies on using machine learning to predict O&M costs are relatively rare. This is probably because obtaining accurate maintenance data is a challenging (Neely & Neathammer, 1991). The most commonly used machine learning methods in O&M costs forecasting are multiple regression (C. S. Li & Guo, 2012a; Au-Yong et al., 2014; Weerasinghe et al., 2016; Krstić & Marenjak, 2017) and ANN (C. S. Li & Guo, 2012a; Tu & Huang, 2013). Au-Yong et al. (2014) found that the characteristics of condition-based maintenance of the office buildings directly influence the cost performance. Based on these relationships, they developed a regression model for maintenance planning and prediction. Krstić & Marenjak (2017) developed a multiple regression model to predict the O&M costs for university buildings during the initial design phase. Li & Guo (2012b, 2012a) developed maintenance cost prediction models for university buildings using simple linear regression, multiple regression, and back-propagation ANN. The

results indicated that the back-propagation ANN model outperforms the other two models. They also found that, for university buildings, the first peak of renovation will be around 20 years of age and second peak 35; for a building with more than five floors, the first and second peak of renovation will be 15 years and 30 years, respectively.

The most commonly used descriptive attributes for developing the O&M costs prediction models involve 1) building age, 2) number of rooms, 3) building floor area, and 4) number of floors.

DISCUSSIONS ON RESEARCH GAPS AND RECOMMENDATIONS FOR FUTURE RESEARCH

Overall Life-cycle Cost Prediction

Although machine learning techniques have been implemented in forecasting construction costs, utility consumption, and O&M costs, its application in predicting a building's life-cycle cost is rarely found in the literature. As one of the few examples, Kang (2017) proposed a multiple regression-based LCC model of the nearly Zero Emission Building for decision making in the early design phase. More studies that utilize machine learning to predict a building's overall life-cycle costs and shed light on the underlying relationships between each cost components (initial costs, maintenance costs, electricity costs, etc.) are needed. With the developments of building systems, increasing amount of facility-related data are being generated, such as the progressively detailed energy consumption data in the Building Automation System (BAS) and maintenance work order history in Computerized maintenance management system (CMMS). With these LCC-related data extracted, cleaned, and stored in one database, and applying machine learning techniques we can forecast each LCC component of a building or a building system, and ultimately predict the whole life-cycle costs.

Generalizable Machine Learning Frameworks for Building LCCA

Most of the developed machine learning models are only applicable to one type of building projects, such as housing (Hong et al., 2011; Jin et al., 2016), educational buildings (C. S. Li & Guo, 2012a), and office buildings (C. W. Koo et al., 2010). The nature of predictive models involves assumptions and simplifications based on the similarities of the studied subjects. The uniqueness of different building projects basically makes it impossible to use one model to predict more than one type of building's LCC (Hong et al., 2011; Bala et al., 2014; Banihashemi et al., 2017). However, it is possible to develop generalizable frameworks of machine learning models for building LCCA. These frameworks would specify the means and process of 1) identifying potential descriptive attributes, such as by literature review and survey, 2) data acquisition, such as by exporting from Building Information Models, BAS, and CMMS, by finding the records in drawings and specifications, and by survey, 3) attributes selection, 4) machine learning algorithm selection, and 5) model validation.

Data Availability, Accessibility, and Quality

Many research challenges discussed can be attributed to data insufficiency, including a lack of sufficient metering and accessibility, and poor data quality (Gallagher et al., 2018). The machine learning models in many studies are established based on a very limited data set. For example, Sonmez (2008) constructed the construction cost model based on the data compiled

from 20 projects. Similarly, Shi & Li (2008) used 14 buildings to build the construction cost prediction model. As Milion et al. (2016) pointed out, “data survey is the most difficult challenge in estimation studies”. Limited and uncertain information makes the accurate prediction of construction-related costs difficult (C. Koo et al., 2011). The lack of reliable and consistent data also limits the application of LCCA in the early design stage (Weerasinghe et al., 2016). In the future, data availability and accessibility issues can be addressed by the developments of Internet of Things (IoT)-related technologies in the building sector. IoT envisions a future in which digital and physical entities can be linked through embedded identification, sensing, and/or actuation capabilities to enable various innovative applications and services that improve the quality of human life (Xia et al., 2012). The built environment is a critical component of the overall IoT network (Gao et al., 2018). Smart Home applications are examples of the new functions in the built environment enabled by IoT technologies. Research on how to improve data availability, accessibility and quality in the building sector through the IoT technologies is greatly needed.

CONCLUSION

Machine learning-enabled facility LCC prediction is a new and growing area of research. Through a literature review, this research summarizes current machine learning applications in the prediction of initial construction costs, utility consumption, and O&M costs. Publications studied in this research have shown that future research is needed 1) to utilize machine learning to predict a building’s overall life-cycle costs and to shed light on the underlying relationships between each LCC components (initial costs, maintenance costs, electricity costs, etc.); 2) to develop generalizable frameworks of developing machine learning models for building LCC analysis; 3) to improve data availability, accessibility, and quality in the building sector through the IoT) technologies.

REFERENCE

- Aibinu, A. A., Dassanayake, D., Chan, T.-K., & Thangaraj, R. (2015). Cost estimation for electric light and power elements during building design: A neural network approach. *Engineering, Construction and Architectural Management*, 22(2), 190-213.
- Alshamrani, O. S. (2017). Construction cost prediction model for conventional and sustainable college buildings in North America. *Journal of Taibah University for Science*, 11(2), 315-323.
- Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable & Sustainable Energy Reviews*, 81, 1192-1205.
- An, S.-H., Kim, G.-H., & Kang, K.-I. (2007). A case-based reasoning cost estimating model using experience by analytic hierarchy process. *Building and Environment*, 42(7), 2573-2579.
- Au-Yong, C. P., Ali, A. S., & Ahmad, F. (2014). Prediction cost maintenance model of office building based on condition-based maintenance. *Maintenance and Reliability*, - 16(- 2), - 324.
- Bala, K., Ahmad Bustani, S., & Shehu Waziri, B. (2014). A computer-based cost prediction model for institutional building projects in Nigeria: an artificial neural network approach. *Journal of Engineering, Design and Technology*, 12(4), 519-530.
- Banihashemi, S., Ding, G., & Wang, J. (2017). Developing a hybrid model of prediction and classification algorithms for building energy consumption. In F. Alam, R. Jazar & H.

- Chowdhury (Eds.), *1st International Conference on Energy and Power, Icep2016* (Vol. 110, pp. 371-376).
- Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. A. (2018). Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm: Comparison with Machine Learning Approaches. *Energies*, 11(7).
- Cheng, M.-Y., Tsai, H.-C., & Sudjono, E. (2010). Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry. *Expert Systems with Applications*, 37(6), 4224-4231.
- Chou, J. S., & Ngo, N. T. (2016). Time series analytics using sliding window metaheuristic optimization-based machine learning system for identifying building energy consumption patterns. *Applied Energy*, 177, 751-770.
- Dogan, S. Z., Arditi, D., & Gunaydin, H. M. (2006). Determining attribute weights in a CBR model for early cost prediction of structural systems. *Journal of Construction Engineering and Management-Asce*, 132(10), 1092-1098.
- Dogan, S. Z., Arditi, D., & Gunaydin, H. M. (2008). Using decision trees for determining attribute weights in a case-based model of early cost prediction. *Journal of Construction Engineering and Management-Asce*, 134(2), 146-152.
- Dursun, O., & Stoy, C. (2016). Conceptual Estimation of Construction Costs Using the Multistep Ahead Approach. *Journal of Construction Engineering and Management*, 142(9).
- Fuller, S. (2010). Life-cycle cost analysis (LCCA). *National Institute of Building Sciences, An Authoritative Source of Innovative Solutions for the Built Environment* ([www. WBDG. org](http://www.WBDG.org), Washington).
- Gallagher, C. V., Leahy, K., O'Donovan, P., Bruton, K., & O'Sullivan, D. T. J. (2018). Development and application of a machine learning supported methodology for measurement and verification (M&V) 2.0. *Energy and Buildings*, 167, 8-22.
- Gao, X., Tang, S., Pishdad-Bozorgi, P., & Shelden, D. (2018). *Foundational Research in Integrated Building Internet of Things (IoT) Data Standards*. Center for the Development and Application of Internet of Things Technologies (CDAIT). Accessed Jan 24, 2019, retrieved from https://cdait.gatech.edu/sites/default/files/georgia_tech_cdait_research_report_on_integrated_building_-_iot_data_standards_september_2018_final.pdf
- Geysen, D., De Somer, O., Johansson, C., Brage, J., & Vanhoudt, D. (2018). Operational thermal load forecasting in district heating networks using machine learning and expert advice. *Energy and Buildings*, 162, 144-153.
- Hong, T., Hyun, C., & Moon, H. (2011). CBR-based cost prediction model-II of the design phase for multi-family housing projects. *Expert Systems with Applications*, 38(3), 2797-2808.
- Idowu, S., Saguna, S., Ahlund, C., & Schelen, O. (2016). Applied machine learning: Forecasting heat load in district heating system. *Energy and Buildings*, 133, 478-488.
- Jafarzadeh, R., Wilkinson, S., Gonzalez, V., Ingham, J. M., & Amiri, G. G. (2014). Predicting Seismic Retrofit Construction Cost for Buildings with Framed Structures Using Multilinear Regression Analysis. *Journal of Construction Engineering and Management*, 140(3).
- Jain, R. K., Smith, K. M., Culligan, P. J., & Taylor, J. E. (2014). Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. *Applied Energy*, 123, 168-178.
- Ji, C., Hong, T., Jeong, K., & Leigh, S. B. (2014). A model for evaluating the environmental

- benefits of elementary school facilities. *Journal of Environmental Management*, 132, 220-229.
- Jin, R., Han, S., Hyun, C., & Cha, Y. (2016). Application of Case-Based Reasoning for Estimating Preliminary Duration of Building Projects. *Journal of Construction Engineering and Management*, 142(2).
- Kang, H. J. (2017). Development of an Nearly Zero Emission Building (nZEB) Life Cycle Cost Assessment Tool for Fast Decision Making in the Early Design Phase. *Energies*, 10(1).
- Kim, G.-H., An, S.-H., & Kang, K.-I. (2004). Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Building and environment*, 39(10), 1235-1242.
- Kim, G.-H., Yoon, J.-E., An, S.-H., Cho, H.-H., & Kang, K.-I. (2004). Neural network model incorporating a genetic algorithm in estimating construction costs. *Building and Environment*, 39(11), 1333-1340.
- Koo, C., Hong, T., & Hyun, C. (2011). The development of a construction cost prediction model with improved prediction capacity using the advanced CBR approach. *Expert Systems with Applications*, 38(7), 8597-8606.
- Koo, C. W., Hong, T., Hyun, C. T., Park, S. H., & Seo, J. o. (2010). A study on the development of a cost model based on the owner's decision making at the early stages of a construction project. *International Journal of Strategic Property Management*, 14(2), 121-137.
- Krstić, H., & Marenjak, S. (2017). Maintenance and operation costs model for university buildings. *Technical Gazette*, - 24, - 200.
- Li, C., Ding, Z. X., Zhao, D. B., Yi, J. Q., & Zhang, G. Q. (2017). Building Energy Consumption Prediction: An Extreme Deep Learning Approach. *Energies*, 10(10).
- Li, C. S., & Guo, S. J. (2012a). Development of a Cost Predicting Model for Maintenance of University Buildings. In F. L. Gaol & Q. V. Nguyen (Eds.), *Proceedings of the 2011 2nd International Congress on Computer Applications and Computational Science, Vol 1* (Vol. 144, pp. 215-221).
- Li, C. S., & Guo, S. J. (2012b). Life cycle cost analysis of maintenance costs and budgets for university buildings in Taiwan. *Journal of Asian Architecture and Building Engineering*, - 11(- 1), - 94.
- Li, H., Shen, Q., & Love, P. E. (2005). Cost modelling of office buildings in Hong Kong: an exploratory study. *Facilities*, 23(9/10), 438-452.
- Lowe, D. J., Emsley, M. W., & Harding, A. (2006). Predicting construction cost using multiple regression techniques. *Journal of construction engineering and management*, 132(7), 750-758.
- Milion, R. N., Paliari, J. C., & Liboni, L. H. B. (2016). Improving consumption estimation of electrical materials in residential building construction. *Automation in Construction*, 72, 93-101.
- Mocanu, E., Nguyen, P. H., Kling, W. L., & Gibescu, M. (2016). Unsupervised energy prediction in a Smart Grid context using reinforcement cross-building transfer learning. *Energy and Buildings*, 116, 646-655.
- Moon, J., Park, J., Hwang, E., & Jun, S. (2018). Forecasting power consumption for higher educational institutions based on machine learning. *Journal of Supercomputing*, 74(8), 3778-3800.
- Neely, E. S., & Neathammer, R. (1991). Life-cycle maintenance costs by facility use. *Journal of Construction Engineering and Management-Asce*, 117(2), 310-320.

- Noshadravan, A., Miller, T. R., & Gregory, J. G. (2017). A Lifecycle Cost Analysis of Residential Buildings Including Natural Hazard Risk. [Article]. *Journal of Construction Engineering and Management*, 143(7), 10.
- Park, B. R., Choi, E. J., Hong, J., Lee, J. H., & Moon, J. W. (2018). Development of an energy cost prediction model for a VRF heating system. *Applied Thermal Engineering*, 140, 476-486.
- Robinson, C., Dilkina, B., Hubbs, J., Zhang, W. W., Guhathakurta, S., Brown, M. A., et al. (2017). Machine learning approaches for estimating commercial building energy consumption. *Applied Energy*, 208, 889-904.
- Sajadfar, N., & Ma, Y. S. (2015). A hybrid cost estimation framework based on feature-oriented data mining approach. *Advanced Engineering Informatics*, 29(3), 633-647.
- Sala-Cardoso, E., Delgado-Prieto, M., Kampouropoulos, K., & Romeral, L. (2018). Activity-aware HVAC power demand forecasting. *Energy and Buildings*, 170, 15-24.
- Shi, H. W., & Li, W. Q. (2008). *The Integrated Methodology of Rough Set Theory and Artificial Neural-Network for Construction Project Cost Prediction*.
- Sonmez, R. (2008). Parametric range estimating of building costs using regression models and bootstrap. *Journal of construction Engineering and Management*, 134(12), 1011-1016.
- Trost, S. M., & Oberlender, G. D. (2003). Predicting accuracy of early cost estimates using factor analysis and multivariate regression. *Journal of Construction Engineering and Management*, 129(2), 198-204.
- Tu, K. J., & Huang, Y. W. (2013). Predicting the operation and maintenance costs of condominium properties in the project planning phase: An artificial neural network approach. *International Journal of Civil Engineering*, 11(4A), 242-250.
- Weerasinghe, A., Ramachandra, T., & Rotimi, J. O. B. (2016). *A Simplified model for predicting running cost of office buildings in Sri Lanka*. Paper presented at the Proceedings of the 32nd Annual ARCOM Conference.
- Xia, F., Yang, L. T., Wang, L., & Vinel, A. (2012). Internet of things. *International Journal of Communication Systems*, 25(9), 1101.
- Zayed, T. M., & Halpin, D. W. (2005). Productivity and cost regression models for pile construction. *Journal of Construction Engineering and Management*, 131(7), 779-789.
- Zhang, C., Cao, L. W., & Romagnoli, A. (2018). On the feature engineering of building energy data mining. *Sustainable Cities and Society*, 39, 508-518.