UNB | Off-site Construction Research Centre
EST. 1785 UNIVERSITY OF NEW BRUNSWICK

CSCE CCGC 1887 Canadian Society for Civil Engineering

Transforming Construction with Reality Capture Technologies: The Digital Reality of Tomorrow

August 23-25, 2022, Fredericton, New Brunswick, Canada

FLEXIBLE DATA MINING FOR PRIORITIZING CRITICAL BUILDING COMPONENTS

Mostafa, Kareem[1,*], Ahmed, Hani[2], and Hegazy, Tarek[1]
[1] Department of Civil and Environmental Engineering, University of Waterloo, Canada
[2] Department of Civil Engineering, King Abdelaziz University, Saudi Arabia
[*] ktmostafa@uwaterloo.ca

Abstract: Prompt diagnosis and resolution of building problems is essential for facility management personnel to maintain an acceptable level of service for their buildings. While the condition of many building components is assessed based on detailed inspection reports, some components depend more on maintenance requests received from occupants. Currently, facility management teams analyze these variety of data and decide on the required rehabilitation strategy, which is a manual process that is inefficient and subjective. This paper demonstrates the flexible use of data mining techniques on two sample building components with different data types, to support asset rehabilitation decisions. The paper focuses on roofing elements as an example of hidden building systems that are assessed by expert inspectors, and HVAC systems which are assessed based mainly on occupants' feedback. The paper discusses how data mining can be applied flexibly to data samples of the two building components. The presented approach aims to streamline the asset prioritization process for large owner organizations such as university campuses and school boards, especially in the case of limited budgets.

Keywords: Facility Management, Capital Renewal, Rehabilitation, Text Mining, Data Mining, Fund-Allocation, Prioritization

1 INTRODUCTION

Typically, building components are kept in good service condition through rigorous maintenance activities (preventive and corrective), as well as predictive upgrades and rehabilitation decisions (Hegazy and Gad 2014). Preventative maintenance activities are conducted every pre-set period regardless of the asset condition. While this strategy ensures minimal failures, it does not benefit from the possible remaining service life of a component. In parallel with preventive actions, corrective (reactive) maintenance actions are also needed, particularly for old buildings, to respond to urgent calls to fix sudden failures and asset repairs. Also, in parallel with these two maintenance activities, building components require well-structured rehabilitation (capital renewal) decisions to upgrade building components, increase capacity and service life, modernize systems (e.g., reduce energy cost), and increase building value.

Asset condition assessments serve as the backbone of an efficient smart rehabilitation planning. Currently, most inspections and condition assessments are performed manually and delivered in the form of hand-written reports. This makes condition assessments a daunting task especially in the case of large asset portfolios such as university campuses and school boards. Furthermore, the submitted reports are often subjective and lack the sufficient granularity to help the facility manager objectively classify the required rehabilitation activities (Ahluwalia and Hegazy 2010). Hence, there have been calls for an objective data-driven approach to enable better fund allocation decisions, similar to the “Moneyball” approach adopted by the Oakland A’s manager Billy Beane where he relied on player data and performance metrics to build a top-tier baseball team using a limited budget (Adriaens 2017).

Supporting asset condition assessments through data mining techniques has been the focus of many recent research efforts. Hegazy and Gad (2014) used two years of reactive maintenance data from 88 schools in Canada to prioritize inspection efforts by identifying systems that had data inconsistencies. Gunay et al. (2019) and Mo et al. (2017) analyzed building work orders and maintenance requests to detect failure patterns and help assign the needed maintenance crews. Similar work was done on bridge inspection reports by Liu and El-Gohary (2017) to extract defect type and severity information. Wen et al. (2019) developed a data management framework combining relational and time-series databases to enhance buildings operation and maintenance. Mostafa et al. (2021) used clustering techniques to classify the roofing elements of over 400 schools that belong to the Toronto District School Board into four main categories depending on the school age and roofing condition. Other applications related to non-building elements include the works of Zhi-gang and Han-bin (2017) on subway equipment and Wang (2021) on aircraft maintenance.

As such, with a focus on Roofing and Heating, Ventilation, and Air Condition (HVAC) equipment, this paper demonstrates the use of data mining to prioritize building assets based on the condition of their components and the types of data available for decisions, in preparation for fund allocation activities. The paper assists owners of large asset portfolios such as school boards and university campuses better maintain the level of service of their assets while abiding by their financial limitations.

2 METHODOLOGY

The work of this paper demonstrates the use of data mining techniques to extract and analyze the vast amount of information present in typical inspection reports and maintenance requests. As such, the methodology adopted in this paper is divided into three main sections as seen in Figure 1. First, the data necessary to perform the condition assessment are extracted from the corpus of submitted reports. This data includes condition description, asset age, size, location, etc. The next step is analyzing this data to gain insights about the type and severity of the damage each building component has sustained. This is performed using rule-based text mining where descriptive text related to the studied damage type are extracted and analyzed. Finally, based on the results of the previous steps, classification and/or clustering is then used to triage the different components into groups based on their condition, thus aiding the asset managers in their prioritization and fund allocation decisions.

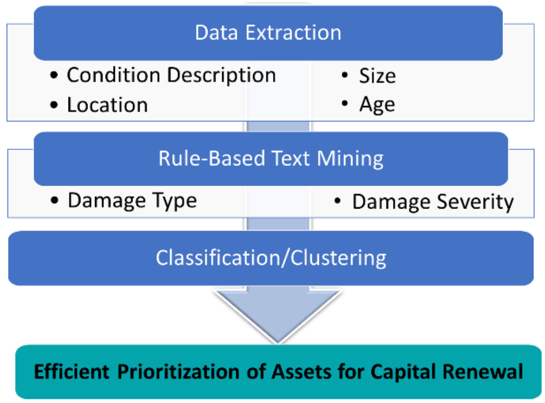

Figure 1: Research Methodology

3 DATA STRUCTURE

To assist with their management analyses and decisions, asset managers break down their assets into a defined hierarchy that encompasses all the building components and aggregates them into subsystems (e.g., interior, exterior, foundation) and then systems (architectural, structural, mechanical). For each component, information is stored that relates to its age, rehabilitation history, and operating condition. This data is typically updated each inspection cycle. An example of such hierarchy is seen in Figure 2. Roofs & HVAC are selected to show two variety of data sources; detailed inspection reports submitted by technical professionals (Roofs), and maintenance requests and complaints submitted by the facility end users (HVAC)

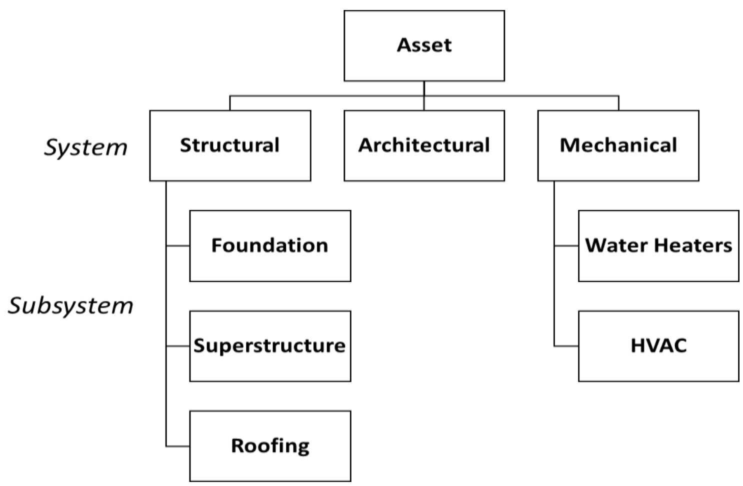

Figure 2: Sample Building Hierarchy

Various studies have shown that roofs are among the building components in most need of repairs (Mostafa et al. 2021, Morgado et al. 2017, Garcez et al. 2012). This is because roofs are frequently exposed to a variety of environmental factors such as extreme temperatures, rain, snow, solar radiation, and others. Built-up roofs (BURs) are among the most common roofing types in Canada (Ahluwalia and Hegazy 2010). Roofing defects can take multiple forms such as ponding and water accumulation, vegetation, leaks, as well as tears and punctures in the roofingmembranes. In addition to the roof condition assessment, a typical inspection report includes information about the building (e.g., age, size). The data from these inspection reports are aggregated into a spreadsheet database for analysis such as the one shown in Figure 3.

AssetID	Constr. Year	Size	Roof Condition
1	1968	15219	The roof has surpassed its theoretical useful life. Roof leaks have been indicated. Replace roof.
2	1,926	4,851	Roofing over gymnasium is original and has history of leaks. Replacement is recommended.
3	1929	19554	The Auditorium and stage roofs (G and H) were replaced a year ago. The remaining roof sections exhibit ponded wate
4	1914	1,681	Existing roof consists of built up and pitch roof. Pitch roof shingles damaged and deteriorated. School representative
5	1981	3,730	The roof is the original roof installed in 1981 at the time of construction. Based on age the roof has surpassed its usef
6	1976	22964	Due to snow and ice cover at the time of the site visit, the condition of the roof could not be verified. It was indicated
7	1958	3,473	Blisters, ridges, and water ponding were observed on the remaining original sections of the roof. Replacement is requi
8	1965	6,346	Localized areas of ponded water and bleed-throughs were observed. The roof has exceeded its theoretical life and re
9	1960	10253	The roof sections on the 'L' shaped roof were replaced approximately 10-15 years ago and currently indicate prematu

Figure 3: Database of Roofing Condition Assessments

In addition to roofing, HVAC equipment is an area of major concern to asset managers. In addition to the importance of HVAC in providing comfort to the building users, especially in places with extreme climates such as Canada and the middle east, HVAC equipment is the most energy consuming service in the building (Gonzales-Torres et al. 2021) representing 32% of the overall building energy usage worldwide. As such, maintain the condition of HVAC equipment is key to building energy efficiency. Typical HVAC problems are either leakage, mechanical issues (e.g., not enough cooling/heating), or unacceptable odors emanating from the ventilation ducts. An example of a database housing the different maintenance requests is shown in Figure 4.

Request ID	Condition	Floor	Building	Reported Date
2934100	AC not working properly in some rooms	D-090	D	30/08/20 10:04:00 ص
3032191	High humidity in bedroom (80%)	A-122	A	12/10/20 10:42:02 ص
3112033	foul odors in master bedroom and AC not working properly	C-058	C	09/12/20 12:06:57 م
3341763	AC not working (too hot)	B-136	B	24/03/21 01:02:23 م
3371339	Loud noise coming from AC	B-147	B	29/03/21 07:51:32 ص
3452025	Loud noise coming from AC	A-098	A	02/05/21 10:08:05 ص
3782969	foul odors from AC	C-027	C	06/09/21 09:21:29 ص
4007212	Smells from neighboring unit leak into this unit	B-170	B	29/11/21 11:26:32 ص
4007521	AC temperature inconsistent	A-082	A	01/12/21 11:15:05 ص
3341901	AC hot, dust comes from AC	F-075	F	25/03/21 12:18:55 م

Figure 4: Database of HVAC Maintenance Requests

4 ANALYSIS

Given the vast amount of textual data available in each report, text mining techniques had to be relied upon to gain better insights at the condition of each building component. First, the textual descriptions in the database was inspected to determine the main keywords that correspond to each defect type. For example, in roofing components, word descriptors such as “surpassed” or “replacement” represent the defect category “aging”. Then a script was developed to search the textual “condition assessment” field in the Roofing database (the right most column in Figure 2) and the “Condition” field in the HVAC database (second column in Figure 3) for the occurrence of those keywords. This serves two purposes; first, it helps define the type of defect (if any) that each component experiences, second, it helps define the severity of those defects through the frequency of the occurrence of those keywords (the higher the frequency, the worse the component condition). The pseudocode used can be seen in Figure 5, where the term “Rehabilitation_Event” refers to each entry in the database. Examples of the aforementioned text extraction process can be seen in Table 1.

```
For each Rehabilitation_Event
  Full_Text = Condition
  Total_Count = 0
  For i = 1 to Defect_Categories
        Defect_Count (i) = 0
        For each Word in Keyword_List (i)
            If Word in Full_Text Then
                Defect_Count (i) = Defect_Count (i) + 1
                Total_Count = Total_Count + 1
             End If
        Next Word
  Next List
  Save Defect_Count (1 to Defect_Categories)
  Save Total_Count
Next Event
```

Figure 5: Text Mining Pseudocode

Table 1: Examples of Text Mining Process

Building Component	Sample Text	Defect Count (occurrence of keywords)	Total Count
Roofing	Original Roof Installed in 1980. Roof **Leaks** have been observed. Roof has **surpassed** its useful life. **Replacement** is recommended	Leaks = 1 Aging = 2	3
HVAC	**odors** in master bedroom and AC **not working** properly	Smell = 1 Mechanical = 1	2

Based on the results of text-mining, classification of the different components according to their conditions can be performed in a quantitative manner. For example, analyzing the HVAC components of a residential compound composed of six different buildings produces the results seen in Table 1. It can be seen that building A is the one in need of most immediate rehabilitation attention, while building E is one with the healthiest HVAC equipment.

Table 2: Classification Results based on Text Mining

Defect Type	**A**	**B**	**C**	**D**	**E**	**F**
Leak	15	10	8	3	3	4
Mech	8	4	3	5		
Smell	15	14	18	10	3	7
Grand Total	**38**	**28**	**29**	**18**	**6**	**11**

Another type of analysis can be conducted using more than just the condition assessment attributes. An example can be seen in Figure 6, combining age attributes and text mining results of the roofing elements for 400 schools, clustering can then be used to triage the different schools according to their age and condition. For example, the most critical buildings (high age and word count) are grouped in the blue cluster, followed by the green cluster (high age and low word count).

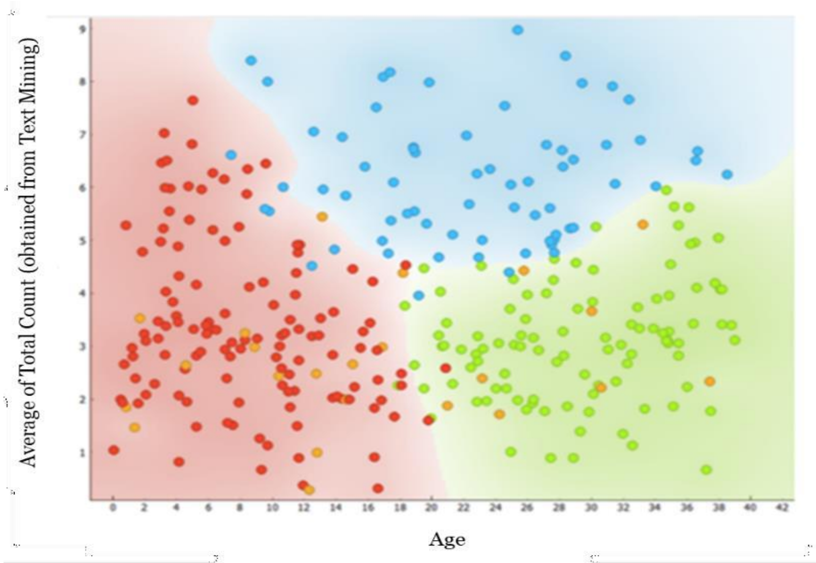

Figure 6: Results of Clustering Analysis Using Age and Text Mining Results

Based on the clustering results, future work would include aggregating more data sources such as image analysis for asset components that exhibit visual defects such as structural components and aggregating the results of different component assessments to provide an overall condition evaluation of the asset as “the sum of its parts”. Furthermore, financial information regarding the estimated repair costs for each component can now be used as an element for a rehabilitation optimization scheme, optimizing the overall improvements gained from repairs (improved asset condition – current asset condition) while meeting the budgetary and other resource constraints.

5 CONCLUSION

Applied to two building components with different condition assessment data, this paper presented the advantages of utilizing data mining techniques to provide a more granular and objective prioritization for asset criticality. The presented approach aims to streamline the asset prioritization process for large owner organizations such as university campuses and school boards, especially in the case of limited budgets. While this work has been demonstrated in this paper using only HVAC and roofing datasets, the proposed approach can be expanded to other asset components and/or asset types. Experienced plant operations professionals, after being presented with the results of this work, appreciated the ability to objectively and promptly prioritize the asset components in most critical condition and apply the necessary rehabilitation funds accordingly. This would help asset management professionals make the best use of their budgets to prolong the service life of their asset portfolio.

6 REFERENCES

Adriaens, J. 2017. 'Moneyball' approach to closing the $2 trillion infrastructure finance gap. The Hill.

Garcez, N., Lopes, N., Brito, J. D., and Silvestre, J. 2012. System of inspection, diagnosis and repair of external claddings of pitched roofs. Construction and Building Materials, 35: 1034– 1044.

González-Torres, M., Pérez-Lombard, L., Coronel, J. F., Maestre, I. R., and Yan, D. 2022. A review on buildings energy information: Trends, end-uses, fuels and drivers. Energy Reports, 8: 626-637.

Gunay, H. B., Shen, W., and Yang, C. 2018. Text-mining building maintenance work orders for component fault frequency. Building Research & Information, 47(5): 518–533.

Hegazy, T. and Gad, W. 2014. Dynamic System for Prioritizing and Accelerating Inspections to Support Capital Renewal of Buildings, Journal of Computing in Civil Engineering, 29(5): 04014058

Liu, K., and El-Gohary, N. 2017. Ontology-based semi-supervised conditional random fields for automated information extraction from bridge inspection reports. Automation in Construction, 81: 313–327.

Mo, Y., Zhao, D., Syal, M., and Aziz, A. 2017. Construction Work Plan Prediction for Facility Management Using Text Mining. Proceedings of the ASCE International Workshop on Computing in Civil Engineering 2017, Seattle, Washington. pp. 92-100

Morgado, J., Flores-Colen, I., Brito, J. D., and Silva, A. 2017. Maintenance Planning of Pitched Roofs in Current Buildings. Journal of Construction Engineering and Management, 143(7): 05017010.

Mostafa, K., Atalla, A., and Hegazy, T. 2021. Data Mining of School Inspection Reports to Identify the Assets with Top Renewal Priority. Journal of Building Engineering, 41: 102404

Wang, J. 2021. Research on Civil Aircraft Direct Maintenance Cost Analysis and Optimization Based on Data Mining. Journal of Physics: Conference Series. 1992: 042075

Wen, Q., Zhang, J. -P., Hu, Z. Z., Xiang, X. -S., and Shi, T. 2019 A Data-Driven Approach to Improve the Operation and Maintenance Management of Large Public Buildings, in IEEE Access, 7: 176127-176140.

Zhi-Gang, M., and Han-Bin, L. 2017. Study on the Maintenance of Subway Equipment Based on Data Mining Techniques. International Conference on Smart City and Systems Engineering (ICSCSE) 2017, Changsha, China. pp. 209-215.