RELEVANT PROPERTIES AND RELATIONSHIPS FOR CONSTRUCTION DEFECT DATA MINING ON BUILDING INFORMATION MODELS

Joyce P. M. Delatorre[1] and Eduardo Toledo Santos[2]

Abstract: A Building Information Model can be considered a construction data source which contains, among others, information about geometry and topology of construction elements.

These information-rich structures can be a particularly good source for data mining systems on their goal of discovering hidden patterns if these are related to the size, shape and/or position of construction elements. That kind of pattern may be useful in execution quality control and productivity, maintenance, and Post Occupancy Evaluation, among other analyses. However, most of this information is in implicit form (e.g., which elements are in front of a given window) and need to be extracted from the model to be used in data mining systems.

Data mining systems need certain properties from construction components as well as to identify some relationships between objects to be able to identify execution problem patterns related to the geometry and topology of these components. This work aims to determine which are the main properties and relationships needed for performing data mining from BIM models augmented with construction defect data.

Execution quality data and after-sales maintenance records from two companies, comprising 26 projects, were collected and analysed. The identified problems were grouped in five classes. From this analysis, the relevant properties and relationships were proposed. This information can be used as basis for the development of a BIM data mining platform able to identify patterns in construction defects.

Keywords: BIM, Building Information Model, data mining, quality control, construction defect.

1 INTRODUCTION

Given today's technological pace, it is increasingly frequent the availability of new systems to collect and store construction-related data. Also, new sources of data, including sensors, RFID tags, bar codes (Soibelman et al. 2008) and, more recently, mobile and IoT-enabled devices (Alam et al. 2016) are all contributing to an intense growth in the generation and collection of data in the construction industry. Also, among those technologies getting momentum is BIM - Building Information Modelling.

Since many years, BIM is no longer considered just a trend, adopted by few pioneers, but a key innovation in the AEC market (Eastman et al. 2011), focusing the design, construction, and operation of facilities. With the advancement of IoT - Internet of Things, it is expected that not only buildings but the information used to design and construct them will change (Khemlani 2014). Certainly, there will be parameters in building information models reserved for storing additional properties of "intelligent" building elements. Each construction element can have built-in smart sensors that analyse its

---

[1]   M.Sc., Escola Politécnica, University of Sao Paulo, Sao Paulo, Brazil, joyce.delatorre@usp.br

[2]   Assistant Professor, Escola Politecnica, University of Sao Paulo, Sao Paulo, Brazil, etoledo@usp.br

performance and will carry not only identification and location data, but also be able to understand the characteristics of their properties, potential signs of damage and decay, communication protocols with other elements, among others. For example, the use of sensors in concrete structures may determine the quality and load capacity of each structural element, creating a broad base of quality information.

For analysis and processing of this large volume of data, techniques such as KDD - Knowledge Discovery in Databases, including Data Mining are being used. According to Silberschatz et al. (2005), KDD techniques are designed to automatically discover statistical rules and patterns from data. Data mining combines knowledge discovery techniques created by artificial intelligence researchers and statistical analysis, with techniques for efficient implementation, allowing its use in databases with large volumes of information.

Recent research studied the application of data mining techniques in construction information. Part of them developed systems and processes for the application of these techniques (Soibelman et al. 2002; Soibelman et al 2008; Kim et al. 2011; Cheng et al. 2015), while others applied techniques for the discovery of patterns in databases with construction information (Zhou et al. 2010; Cheng and Leu 2011; Chi et al. 2012).

There are a wide variety of data sources that require the design of storage and analysis systems for more structured data than those needed for simple data, such as specifications, change orders, 2D and 3D drawings, images, video and audio files, among others. The high number of characteristics and dimensionality of data makes it inefficient to store and manage data in transactional databases. In addition, the transfer of data to simple data tables or spreadsheets can lead to loss of important information contained in original data formats, such as texture information present in images (Soibelman et al. 2008). According to Liu et al (2015), data mining techniques are not often used in conjunction with Building Information Modelling, although this application could provide important information from the implicit patterns in BIM models.

According to Gu and London (2010), BIM involves the implementation and maintenance of an integrated digital model of all construction information at different stages of the project life cycle in the form of a data repository, including geometric and non-geometric information. For Wülfing et al. (2014), we can consider that it is possible to store in BIM models construction information generated throughout the building life cycle, avoiding the use of different formats of records such as paper documents, tables, texts, reports, among others and providing a unified comprehensive information resource. The storage and use of this information in BIM models, however, is not considered a trivial task yet (Wülfing et al. 2014). The query or retrieval of required information from BIM models is complex, since they do not necessarily contain all the information integrated in a single model due to differences in stakeholder's interests (Mazairac and Beetz 2013) and may be linked to other information resources related to BIM as, for example, cost and time models.

According to Nepal et al. (2012), although the rapid development of BIM has opened numerous opportunities for design and construction and there is a growing number of builders using BIM as a construction management support, there are still numerous challenges related to information extraction from BIM models, which limits the usability of the building models and other subsequent processes. For Mazairac and Beetz (2013) the ability to obtain the necessary information in time is a key success factor in the construction industry. The complexity and lack of structure in the registry of the information generated during execution and use of facilities are some of the causes for difficulty for timely analysing and obtaining the necessary information.

Construction defects are one of the most common problems in construction projects and can have significant costs, both direct and indirect (Aljassmi and Han 2013). Lee et al (2016) analyse the case of defect records. They cite, for example, that the use of defect data, in practice, is very low due to: lack of formal structure in the representation of data, insufficient data input, just text-based search available and insufficient sharing of data. These unstructured data do not follow a machine-readable format, which makes difficult the access, analysis and search for the desired information. Cheng and Leu (2011) highlighted the importance of organizing and classifying common defects in order to turn it into useful information so that participants in the project can understand critical defects and prevent their recurrence.

Fulfilling the specific needs and work environments of each user type requires the creation of query filters, appropriate methods and tools to access BIM information (Wülfing et al. 2014).

Based on presented issues, we highlight the importance of developing mechanisms for analysing construction defect data from BIM models.

Accordingly, the main goal of this research in progress is to extract data from one or more Building Information Models augmented with execution quality control data and, using data mining algorithms, determine latent causes of construction problems detected either during the execution or in the operation phases.

In this work, we report on the identification of the important properties of and relationships between construction elements that needs to be extracted from BIM models for serving as input to downstream data mining processes (Delatorre et al. 2016).

2   RESEARCH METHOD

The research method adopted for this work started a literature review on BIM concepts and data mining. Data collection was performed on the records and documents of two major contractors of the Brazilian market.

Further data were obtained through semi-structured interviews with their execution quality control personnel working on construction sites. The interviews focused on quality control issues, procedures, data collection operations, storage and processing of collected data and the main problems found on site.

In both companies, it was possible to obtain documents related to their quality control processes, like:

- Service Assessment Forms, used for gathering data on execution quality;

- Technical assistance reports from 26 projects, containing issues and their likely causes;

- Crack maps (a graphical report sent to structural consultants for analysis to determine the probable causes of structure problems and the remediation to be performed on site).

The information from service assessment forms, technical assistance reports and lessons-learned databases from the contractors studied in this research were organized and listed in a spreadsheet. Columns on this spreadsheet registered the problem, its identified cause, and the related building system. It was used as a tool for identifying the main patterns of construction problems, along with their causes, that could be the target of a data mining system focused on quality flaws.

The description of each reported problem was analysed and categorized in groups based on their cause.

As a final step, after analysing the resulting groups and elements, as well as the technical literature, the relevant properties and relationships to be considered in a BIM data mining system for construction defects were derived.

3   DATA COLLECTION

The data collection procedures reported here were developed with professionals directly involved with the quality control processes of the studied companies. Company A is part of a 76-year old holding operating in 20 out of the 26 states in Brazil, with 58,000+ employees. It works in EPC contracts with verticals in energy, transportation, shipbuilding, and land development. Company B was founded in 1973 and has built more than 7 million square meters, operating all around Brazil. It has more than 700 employees and focuses on commercial, residential and industrial buildings, hotels, hospitals and the Oil & Gas sector.

3.1   Company A: digital quality control and non-conformance management processes.

Company A has an integrated non-conformance management system that uses BIM and is based on the Autodesk BIM 360 Field mobile software. Both the manager responsible for BIM implementation and the one responsible for the development of this management system were interviewed, as well as the construction quality control supervisor.

The quality control process based on the mentioned mobile application uses forms which are filled out on tablets and cell phones. They contain the items to be verified on each system. During the assessment visit, all items are verified and, if a non-conformance is detected, general data about the problem is registered in the system and a photo is taken. These registers from the field feed a general database with the following information: problem description; status; form number; element verified; action to be taken.

As part of the quality control process, a non-conformance is monitored from the moment it is registered until it is closed. Only then the payment to the involved service provider is authorized.

Autodesk BIM 360 Field allows the use of a BIM model for visualization and registration of a non-conformance by fixing a pin on a 3D view of the building, but no real link is created between the non-conformance record and the model element as it is only of a visual character.

3.2   Company B: manual quality control and non-conformance management processes.

On Company B, quality control and non-conformance management are performed using printed forms called Service Assessment Forms. These forms contain the items that should be checked on each execution phase including its quality criteria, locations to be checked and tolerances.

According to their Technology Centre supervisor, responsible for creating the forms and monitoring their use on the field, the field crew often fills the forms, but do not input their data on the system, limiting data querying. Only in cases where a severe problem is detected, the forms are consulted for tracking possible causes. This happens, as stated by a quality field engineer, due to the difficulty in tabulating all the data on the filled forms, as they are in an analog (paper) format.

After the building is finished, the quality process goes under the Technical Assistance Department, responsible for analysing and handling all complaints from customers regarding construction problems identified after sales. For each complaint, the following

data is registered: date; building; problem description; involved technologies and systems; and status. This database is used both for managing the complaints and for analysing recurring problems providing construction process feedback (lessons learned).

3.3 Results

Tabulation of the data from service assessment forms, technical assistance reports and lessons learned supplied by the studied companies resulted in a table with **425 different problem descriptions** from **26 individual projects** in the period 2013 to 2015. Those descriptions were analysed and, based on their cause, classified into the following categories:

- Element properties: problems related to design specification and materials;
- Execution date: problems related to a specific execution date (e.g.: when some anomalous fact affected labour; or construction before a specific design change; or severe weather on a certain date; etc.) or detection of problems with the batch of the material used.
- Positioning: problems that may be influenced by the position of the element in the building;
- Element relationship: problems caused by a relationship between the affected element and other building elements;
- Element geometry: problems related to the element shape;
- N/A: for descriptions where problem cause is not clearly stated or problems caused by accident or failure during execution.

These categories were created during the problem analysis step. Each problem was checked against the proposed category and, if it could not be fitted on an existing one, a new category was proposed.

Most problems were classified into the "Element relationship" class (68.7%), followed by "N/A" (11.5%), "Positioning" (10.4%) and "Element properties" (6.6%).

4 RELEVANT PROPERTIES AND RELATIONSHIPS FOR DATA MINING DEFECT DATA ON BIM MODELS

Based on the categories identified from the collected data, and with support of the literature review for identifying the cause of problems, the set of properties and relationships on Table 1 were considered relevant for extraction from BIM models to be analysed by data mining algorithms in search of patterns from defect data. For each property or relationship, Table 1 contains an example of a possible data query involving the property/relationship and a problem cause that could be uncovered from it.

The few properties listed on Table 1 (material-related properties, execution date and, position), together with the listed relationships (directional, boolean and topological relationships, distance and, geometry) can represent the cause of all the 425 problems identified in the 26 projects analysed in this study. They are very representative for the project class studied (low to medium-rise residential and commercial buildings).

Therefore, these properties and relationships are proposed as suitable data to be extracted from Building Information Models and fed in data mining pipelines in the search for hidden patterns that could explain not obvious causes of problems in buildings.

Table 1: Relevant properties and relationships for mining defect data from BIM

Property / Relationship	Example data query	Example problem cause
Material properties (design specs)	Affected columns vs poured concrete specs	Flawed columns used concrete with fck inferior to the specified in design
Execution date	Affected structural elements vs date of disassembly of formwork / shoring	Early disassembly of formwork or shoring (before complete cure) resulted in structure deformation or cracking
Positioning (X,Y, Z data)	Affected elements vs position in the building	All affected elements are located at the west façade, more exposed to sun
Directional relationship between elements (above; below; northOf; southOf; eastOf; westOf)	Affected columns vs position relative to affected foundations	Cracks on columns are due to movements of foundations because of soil displacement
Distance between elements (shorter distance between element's bounding boxes)	Affected columns vs distance to building core	Columns located up to 3m from building core present cracks
Boolean relationship between elements (union; intersection; difference)	Affected slabs vs intersection with supporting element	Slabs supported by column #43 are cracked
Topological relationship between elements (touch; contain; equal; inside)	Defects x topologic relation (inside) with affected structural elements	Bean that supports slabs presents concrete voids (defects) along its length
Geometric property	Affected columns x slenderness (width / height ratio)	Problems caused by the design of too much slender structures which, due to the great buckling length, smashed dividing walls

5   CONCLUSION

Based on the interviews and examination of the quality control processes and non-conformance management practices of the studied companies, it was possible to verify the complexity involved in collecting and analysing construction defect data from building sites.

As indicated by the literature review, gathering of these data is usually unstructured, making it difficult to consolidate information in a single database that can be queried for

analysing possible problem causes to provide feedback to the design and construction processes.

The Building Information Model has the potential to became an information repository for all phases of the building lifecycle, including data generated after design, as are those collected during execution quality control.

Although company A is in a more advanced stage, as it collects defect data using mobile devices that feed them to a digital database suitable for analysis, these data are not linked back to the BIM component they refer to. This makes it difficult to identify problem causes that are related to geometry and relationships among construction elements, which are implicitly coded in BIM models.

However, the use of Building Information Models as input for data mining systems is not a trivial task. Data mining techniques allow analysis and processing of large amounts of information in the search for useful patterns, but it is necessary, during data pre-processing, to develop suitable representations for both geometry related properties and relationships among elements so that this important information can be considered during the mining process.

This paper investigated data about construction problems from two companies, seeking to identify the important properties and relationships to be fed in data mining systems designed to work on BIM Models augmented with construction quality data. Three properties and five classes of relationships were identified as sufficient to describe the causes of all the 425 problems reported in the 26 projects analysed in this study.

We believe the properties and relationships identified in this study can also be useful in data mining systems for uncovering patterns related to productivity and facility management, among others, if input is augmented with corresponding information, because they too can be influenced by the geometry and position of construction elements.

The main limitations of this work are: i. the lack of an actual validation within a BIM data mining platform, because such system is not yet available and, ii. the scope of the studied data is limited to building construction and probably is not valid for infrastructures like roads, tunnels, dams, and bridges.

Further work, already in progress, is on construction defect representation in BIM models and proper BIM data representation format for input to standard data mining systems.

6 ACKNOWLEDGMENTS

7 REFERENCES

Alam, F. et al. (2016). Analysis of Eight Data Mining Algorithms for Smarter Internet of Things (IoT). *Procedia Computer Science,* 98(2016), pp. 437-442, doi: 10.1016/j.procs.2016.09.068

Aljassmi, H. and Han, S. (2013). Analysis of Causes of Construction Defects Using Fault Trees and Risk Importance Measures. *ASCE J. Constr. Eng. Manage.*, 139(7), pp. 870-880, doi: 10.1061/(ASCE)CO.1943-7862.0000653

Cheng, Y. and Leu, S. (2011). Integrating data mining with KJ method to classify bridge construction defects. *Expert Systems with Applications*, 38(6), pp. 7143–7150, doi: 10.1016/j.eswa.2010.12.047

Cheng, Y., Yu, W. and Li, Q. (2015) GA-Based Multi-Level Association Rule Mining Approach for Defect Analysis in the Construction Industry. Automation in Construction, 51, pp. 78–91, doi: 10.1016/j.autcon.2014.12.016

Chi, S., Suk, S., Kang, Y. and Mulva, S. P. (2012). Development of a Data Mining-based Analysis Framework for Multi-attribute Construction Project Information. *Advanced Engineering Informatics*. 26(3), pp. 574–581, doi: 10.1016/j.aei.2012.03.005

Delatorre, J.P.M., Corrêa, F.B. and Santos, E. T. (2016). Mining BIM Models: Data Representation and Clustering from Implicit Relationships. 16th International Conference on Computing in Civil and Building Engineering, ICCCBE 2016, Osaka, Japan, pp. 983-990.

Eastman, C. M., Teicholz, P., Sacks, R. and Liston, K. (2011). *BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers and Contractors*, 2nd edn, Wiley, Hoboken, New Jersey.

Gu, N., and London, K. (2010) Understanding and Facilitating BIM Adoption in the AEC Industry, *Automation in Construction*, 19(8), pp. 988-999, doi: 10.1016/j.autcon.2010.09.002

Khemlani, L. (2014). *The "Internet of Things"*. Available at: http://www.aecbytes.com/buildingthefuture/2014/InternetofThings.html [Accessed 02 feb. 2015].

Kim, H., Stumpf, A. and Kim, W. (2011). Analysis of an energy efficient building design through data mining approach. *Automation in Construction*. 20(1), pp. 37–43, doi: 10.1016/j.autcon.2010.07.006

Lee, D., Chi, H., Wang, J., Wang, X. and Park, C. (2016) A Linked Data System Framework for Sharing Construction Defect Information Using Ontologies and BIM Environments, *Automation in Construction*, 68(2016), pp.102-113, doi: 10.1016/j.autcon.2016.05.003

Liu, Y., Li, H., Li. H., Pauwels, P. and Beetz, J. (2015). Recent Advances on Building Information Modeling. *The Scientific World Journal* (2015), Article ID 786598, pp.1-2, doi: 10.1155/2015/786598

Mazairac, W. and Beetz, J. (2013). BIMQL – An open query language for building information models. *Advanced Engineering Informatics*, 27(4), pp. 444–456, doi: 10.1016/j.aei.2013.06.001

Nepal, M. P., Staub-French, S., Pottinger, R. and Webster, A. (2012). Querying a Building Information Model for Construction-Specific Spatial Information. *Advanced Engineering Informatics*, 26(4), pp. 904–923, doi: 10.1016/j.aei.2012.08.003

Silberschatz, A., Korth, H.F., and Sudarshan, S. (2005). *Database System Concepts*, 5th edn, McGraw-Hill, New York.

Soibelman, L., and Kim, H. (2002). Data Preparation Process for Construction Knowledge Generation through Knowledge Discovery in Databases. *ASCE J. Computing in Civil Engineering*, 16(1), pp. 39-48, doi: 10.1061/(ASCE)0887-3801(2002)16:1(39)

Soibelman, L., Wu, J., Caldas, C., Brilakis, I. and Lin, K. (2008). Management and Analysis of Unstructured Construction Data Types. *Advanced Engineering Informatics*, 22(1), pp. 15–2, doi:10.1016/j.aei.2007.08.011.

Wülfing, A., Windisch, R. and Scherer, R. J. (2014). A Visual BIM Query Language. In: A. Mahdavi, B. Martens & R. Scherer, (eds), eWork and eBusiness in Architecture, Engineering and Construction: ECPPM 2014, pp. 157-164. CRC Press, Boca Raton.

Zhou, G., Wang, L., Wang, D. and Reichle, S. (2010). Integration of GIS and Data Mining Technology to Enhance the Pavement Management Decision Making. *Journal of Transportation Engineering*. 136(4), pp. 332–341.