Automation in Construction 91 (2018) 256–272

Contents lists available at ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon

Comparing machine learning and rule-based inferencing for semantic enrichment of BIM models

Tanya Bloch*, Rafael Sacks

Faculty of Civil and Environmental Engineering, Technion, Israel

ARTICLE INFO

Keywords:
Building information modeling
Semantic enrichment
Machine learning

ABSTRACT

The need for extensive pre-processing to prepare model data for sophisticated BIM applications, such as automated code compliance checking, functional simulation and analysis, and information exchange, is a major obstacle to their use. Semantic enrichment offers an alternative, automated approach to manual pre-processing. We illustrate the use of machine learning algorithms for semantic enrichment of BIM models, and compare it to rule-inferencing, through application to the problem of classification of room types in residential apartments. The results showed that machine learning is directly applicable to the space classification problem. Although rule-inferencing has succeeded in other contexts, it proved to be unsuitable for this problem. This leads to the observation that different BIM object classification problems require different approaches. More generally speaking, the AI methods used for semantic enrichment depend on the nature of the problem context, and future research will be needed to establish suitable guidelines.

1. Introduction

In theory, BIM technology and the adoption of the Industry Foundation Class (IFC) standard support the automation of processes like code compliance checking [1], functional simulations, safety planning and design analyses [2,3]. In practice, however, technical difficulties, common working practices and interoperability issues remain that limit the degree of automation that can be achieved [4]. Every analysis tool requires specific information to be present in the BIM model, which usually leads to the need to export multiple models, each one specifically tailored for a specific analysis tool [5,6]. This is wasteful in terms of the resources of BIM vendors, and in many cases needed export formats are unavailable. The challenge to support different analysis and building evaluations based on the same model remains to be achieved.

The major problem with the existing process is to obtain the correct data in the correct representation as required by each interface. Reviews of existing applications for code compliance [1,7] report that while some interfaces for automatic code compliance are available, the phase of retrieving information from the building model is often a source of error due to inaccurate, false information, or lack of information provided by the user in the design phase. Consequently, every iteration of an analysis of a given design requires the user to provide additional information or specify again information that was already given in the original BIM model. Semantic enrichment offers an alternative approach that aims to automate the interface [5]. A successful semantic enrichment tool would infer any missing information required by the receiving application automatically, thus alleviating the need for the user to engage in time-consuming pre-processing of the building model.

Semantic enrichment encompasses classification of building objects, aggregation and grouping, unique identification, completion of missing objects, and reconstruction of occluded objects in the case of application to models compiled from point cloud data (PCD) [8]. BIM objects derive many of their properties from their class, making object classification crucial for reuse in different analysis tools. Classification of spaces, for example, is crucial for spatial validation of a BIM model and for many other analyses. Aggregation and grouping are essential for operations such as quantity take-off and cost estimation. Unique identification, including numbering of objects, is required for tracking objects and reporting results in almost every use case. Completion and/or reconstruction of missing objects or objects with missing parts can often be done if the function of the parts can be determined. Different BIM applications have different native models, each of which classifies, aggregates, identifies and parametrizes its objects differently. Thus semantic enrichment can also help achieve interoperability by enriching models automatically according to the model view definition (MVD) of the receiving application.

Semantic enrichment of BIM models per se is a relatively new area of research and the literature on the subject is limited. However, much

* Corresponding author at: Technion IIT, The National Building Research Institute, 3200003 Haifa, Israel.
E-mail addresses: bloch@campus.technion.ac.il (T. Bloch), cvsacks@technion.ac.il (R. Sacks).

https://doi.org/10.1016/j.autcon.2018.03.018
Received 26 July 2017; Received in revised form 31 December 2017; Accepted 8 March 2018

work has been done towards intelligent semantic query of BIM models [9–11]. These efforts have begun to exploit the meaningful topological constructs of information that is implied in BIM models, but not stored explicitly. They use the implicit meaning, but no information is added to enrich the model during their operations.

The general purpose of this work was to identify those characteristics of models that are candidates for semantic enrichment that determine which possible alternative approaches are most applicable. Specifically, we tested the possibility of using machine learning algorithms for semantic enrichment, and compared it to rule-based inferencing, which has previously been shown to be useful for classification of model objects and other enrichment tasks [5,8].

A detailed example of room type classification based on space function was selected to explore these approaches and the problem characteristics that determine their applicability. An example of the use of space classifications is to check for compliance with user requirements and with building codes. These type of checks exist in commercial software like Solibri Model Checker (SMC) [12] but they have three a priori requirements: a) that space elements explicitly exist in the BIM model, b) that each space element is labeled, and c) that the labels are recognized by the checking software (i.e. drawn from a predefined dictionary of terms). Unfortunately, none of these requirements can be assumed in BIM models in standard practice [13], whether due to incomplete or inaccurate modeling or naming by the modeler, to incomplete export of data from the authoring application, or due to fundamental differences in space function naming conventions. As a result, commercial software requires that users define the space boundaries, aggregate spaces, and apply the correct labels manually every time a model is checked.

The first section of this paper provides background on semantic enrichment in general, and specifically on the problem of object classification and room type classification. We also briefly discuss the connection of BIM to artificial intelligence, elucidating what is lacking from BIM technology to provide "intelligent" models. Section 3 defines the research methodology, and Section 4 provides a definition of a standardized room classification problem that can also be used by other researchers to test their own methods. Section 5 deals with the suggested Machine Learning approach and Section 6 with the rule-based approach. Each section first describes the method and then the results of experiments. Section 7 discusses the results obtained, the differences identified between the two approaches, the limitations of this research, possible applications of the approach and further work. The conclusions are given in Section 8.

2. Background

2.1. Approaches to semantic enrichment of BIM models

Adoption of Building Information Modeling (BIM) in the architecture, engineering and construction (AEC) industry has brought many benefits, but it has also introduced the problem of interoperability between different stakeholders and their BIM platforms [3,14]. The Industry Foundation Classes (IFC) schema was created by Building Smart (previously called the International Alliance for Interoperability) [15] to overcome this problem and enable BIM technology to reach its full potential in contribution to the AEC industry. IFC is a rich product model and it is considered the common format for data exchange in the AEC industry, but it is very complex, providing representations of many entities from different AEC domains [16]. IFC format is meant to provide an opportunity to exchange information between different platforms supporting IFC [17], yet in practice many problems exist when trying to reuse an IFC model file since information can be lost or distorted during the exchange [18]. Although the IFC schema is rich and broad in its coverage, some applications, like code compliance checking and building analysis application, require higher levels of semantics. Solihin et al. [19] proposed to deal with this problem by creating an external platform called "FORNAX" that encapsulates the semantics of building components with emphasis on attributes needed for code compliance checks. Different operations can then be applied to these enriched objects.

Pauwels et al. [20] suggest the adoption of semantic web technology instead of the IFC standard to improve interoperability since it allows interrelation of all kinds of information, including semantics, into one semantic web. The information from the web can be reused and it also allows derivation of implicit information by a reasoning engine based on logical rules.

Semantic enrichment is a process in which domain expert knowledge is used to infer the semantics of a given model. The inferred information can then be added to the model, thereby enriching it and facilitating its use for any given receiving application (i.e. facilitating its interoperability Belsky et al. [5]. proposed a Semantic Enrichment Engine for BIM (SEEBIM) to implement a rule-based approach for semantic enrichment. The engine reads an IFC file, uses a set of inference rules based on expert knowledge to add new facts to its database, then it writes a new IFC file which includes the inferred information [21].

2.2. Artificial intelligence and building information modeling

According to Galle [22], an ideal computer system for building modeling should be integrated, intelligent, and compliant. He defined intelligent models as models with an "ability to maintain semantic integrity". Moving from "ignorant" CAD drawings to BIM achieved most of the goals Galle set for an ideal computer system for building modeling, but that does not mean that BIM models are "intelligent". BIM models consist of elements that are "aware" of their properties and their relationships to other elements and one might argue that these are to some extent intelligent models. For example, if one was to move a column in a model, any beam supported by the column should extend or shorten accordingly. On the other hand, using the same example we can also argue that models remain "ignorant". Systems allow users to take actions that are obviously not rational in the context of building design, so for instance a user can move a column a very large distance and the supported beam will still extend to reach it. This rationality of actions is still missing from BIM models.

"Intelligence" in existing BIM platforms is expressed in the parametric modeling and design intent behavior that maintains design integrity [23]. The artificial intelligence literature however, maintains that making a machine truly "intelligent", requires embedding not only knowledge about the physical world, but with what is referred to as "common sense knowledge" which is the general knowledge about the world [24–26]. This type of knowledge is natural to a human being. It is used for tasks like language processing and computer vision, but it is mostly missing from BIM platforms. For example, if a modeler neglects to place a railing on a balcony, no objection or warning will be raised, unless a purpose-built rule-checking routine is invoked (for example, see [27]). Yet to a human expert the error is immediately obvious. If the width of the opening to the laundry room is not large enough for a washing machine to be carried through it, we understand that this is a design fault, though no red flags will be raised by the BIM platform. As object-oriented systems, BIM tools include representation of objects, their properties, and the relationships between objects, but we are still far from being able to refer to BIM tools as being intelligent.

[28] organized existing definitions of Artificial intelligence (AI) into four approaches, two of which deal with the thought process and two with behavior. The parametric modeling that underlies BIM software tries to deal with the behavioral aspect of elements during design. Semantic enrichment is better associated with the thought process since we use existing information to automatically draw new conclusions.

The SeeBIM semantic enrichment application uses rule inferencing to obtain new facts about the model [21], but rule-inferencing is a particular and limited implementation of AI. One of the definitions of AI which falls into the category of "systems that think like humans" is