Automation in Construction 74 (2017) 39–54

Contents lists available at ScienceDirect

Automation in Construction

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

Construction Safety Clash Detection: Identifying Safety Incompatibilities among Fundamental Attributes using Data Mining

Antoine J.-P. Tixier [a], Matthew R. Hallowell [b,*], Balaji Rajagopalan [b], Dean Bowman [c]

[a] Computer Science Laboratory, École Polytechnique, Palaiseau, France
[b] Civil Engineering Department, University of Colorado at Boulder, CO, USA
[c] Bentley Systems, USA

ARTICLE INFO

Article history:
Received 12 October 2015
Received in revised form 28 October 2016
Accepted 2 November 2016
Available online xxxx

Keywords:
Building Information Modeling
BIM
Advanced Work Packaging
AWP
Safety
Machine learning
Network analysis
Risk
Prevention through design
Information Technology

ABSTRACT

Construction still accounts for a disproportionate number of injuries, inducing consequent socioeconomic impacts. Despite recent attempts to improve construction safety by harnessing emerging technologies and intelligent systems, most frameworks still consider tasks and activities in isolation and use secondary, aggregated, or subjective data that prevent their widespread adoption. To address these limitations, we used a newly introduced conceptual framework and accompanying natural language processing system to extract standard information in the form of fundamental attributes from a set of 5298 raw accident reports. We then applied state-of-the-art data mining techniques to discover attribute combinations that contribute to injuries. We refer to these incompatibilities as "construction safety clashes". The main contribution of our study lies in the methodological advancements that it brings to the construction safety domain. In light of the results obtained, our approach shows great promise to become a standard way of extracting valuable, actionable insights from injury reports in a fully unsupervised way. The use of our methodology could enable construction practitioners to ground their safety-related decisions on objective, empirical data, rather than on limited personal experience or expert opinion, which is the current industry standard. Finally, our methodology allows construction accidents to be viewed as perturbations in underlying networks of fundamental attributes. While the analysis of the current data set provides preliminary evidence for this theory, comparison to non-accident reports will be required for validation.

1. Introduction and Motivation

Even though safety performance has notably improved after the inception of the Occupational Safety and Health Act (OSHA) of 1970, construction fatalities, disablements, and illnesses still have a dramatic socioeconomic impact. In fact, construction still accounts for a fatal occupational injury rate of 9.4 per 100,000 full-time workers, one of the highest in the United States [10]. Moreover, the construction industry has consistently accounted for the most fatalities of any industry in the private sector since 2005, with 796 casualties in 2013 alone. Therefore, improving safety has become an absolute priority.

Construction has reached saturation with respect to the traditional safety strategies that were originally implemented to comply with regulations [25]. Therefore, safety researchers and professionals have recently tried to harness emerging technologies and intelligent systems that are traditionally used for design, planning, or operations. Some examples of such technologies include Building Information Modeling (BIM), proximity sensing, or information retrieval. While these efforts are worthy, they currently suffer limitations, as the data used are mostly secondary, aggregated, and subjective (based on regulations, intuition, or judgment), and tasks are considered in isolation, preventing the efficient capture of the transient and dynamic nature of construction work [65].

To improve the robustness of safety analyses, Esmaeili and Hallowell [26,27] and Esmaeili [23] introduced a conceptual framework where any injury can be characterized by a unique combination of universal context-free descriptors of the work environment, also called fundamental attributes or injury precursors. These works made great strides by showing possible the extraction of objective, standardized structured information from unstructured injury reports, opening the gate for the first time to leveraging big, empirical, and objective safety-related data. However, several major limitations remained, such as the needs for a more comprehensive set of attributes and for an automated system to scan the reports. Prades Villanova [65] and Desvignes [21] addressed the first limitation by proposing a refined and expanded list of fundamental attributes, and Tixier et al. [77] addressed the second by developing a highly accurate (96% in $F1$ score) natural language processing (NLP) system.

In this study, we tested the extent to which graph mining and hierarchical clustering can be used to identify safety-critical associations of attributes from large data sets. We conducted our experiments on an

* Corresponding author.
E-mail address: matthew.hallowell@colorado.edu (M.R. Hallowell).

http://dx.doi.org/10.1016/j.autcon.2016.11.001
0926-5805/

attribute data set obtained from scanning 5298 raw injury reports with Tixier et al.'s [77] NLP system.

2. Background and Point of Departure

This study was built upon a foundation of knowledge in two key areas: construction safety analysis, and safety integration with BIM. Although both of these areas have received some attention from the scientific and practical communities, researchers have yet to explore their nexus. The following literature review highlights current limitations in both domains and develops a firm point of departure.

2.1. Construction safety analysis

Safety analysis in construction has taken many forms and varies greatly in the *data sources* used and the *level of detail of the units of analysis (data granularity)*.

2.1.1. Data sources

The vast majority of construction safety studies rely on opinion-based risk data, generally obtained by asking experts to rate the relative magnitude of risk based on their professional experience and intuition [65]. Such data are subjective and suffer the numerous biases that affect human judgment under uncertainty, such as overconfidence, anchoring, availability, representativeness, unrecognized limits, or conservatism [11,68,78]. Additionally, there is evidence that gender [37] and even emotional state [76] impact risk perception. Although one can attempt to minimize the effects of some of these psychological biases [39], opinion-based data remain severely limited in comparison to empirical data. Therefore, the needs to leverage objective raw empirical data are pressing.

2.1.2. Level of detail of the units of analysis

Construction work is very complex from both technological and organizational perspectives. Even though the multifactorial nature of safety risk is well known [41,70], most studies have decomposed construction processes into smaller parts for the sake of simplicity [58]. Such breakdown allows researchers to model safety for a variety of units of analysis. For example, Hallowell and Gambatese [40] focused on specific worker motions and activities needed for formwork construction, Navon and Kolton [60] analyzed interactions among planned tasks at height, and Huang and Hinze [45] modeled task, location, time, human error, and age as risk factors. Trades have most commonly been adopted as the granularity level [4,28,49]. A limitation of these segmented approaches that consider elements in isolation is that there are a virtually infinite number of units of analysis that must be taken into account in order to comprehensively capture safety. This has prevented the adoption of a robust, standardized way of approaching safety analysis in construction.

2.1.3. Attribute-based approach to construction safety analysis

The attribute-based framework for construction safety was introduced by Esmaeili and Hallowell [26,27] and Esmaeili [23] in an effort to jointly address the data subjectivity and study segmentation limitations previously described. Indeed, this unified approach allows the extraction of standardized safety information from objective, raw textual data such as injury reports. Fundamental attributes are universal, context-free descriptors of the jobsite. They span construction means and methods, environmental conditions, and human factors.

To illustrate, in the following report excerpt: “employee tripped on an electrical cord while exiting job trailer”, three fundamental attributes can be identified: (1) *object on the floor*, (2) *exiting/transitioning*, and (3) *job trailer*.

While simple, this approach is powerful, as any incident can be viewed as the resulting outcome of the joint occurrence of some fundamental attributes and the presence of a worker. It follows that the same standard safety information can be extracted for any construction situation regardless of the trade, task, industry sector, or part of the world in which the accident occurred.

Esmaeili and Hallowell [26,27] initially proposed short lists of fundamental attributes (14 and 34, respectively) identified from analyzing 105 fall and 300 struck-by high severity injury cases drawn from national databases. Prades Villanova [65] and Desvignes [21] refined and broadened these drafts to a final, robust list of 80 carefully engineered and validated attributes by manually analyzing a larger database of 2201 injury reports featuring all injury types and severity levels. These precursors are summarized in Table 1.

However, while the attribute-based framework is particularly well-suited for leveraging big textual safety-related data, the high cost and numerous limitations of manual content analysis remained as serious obstacles to its large-scale implementation. To solve this problem, Tixier et al. [77] developed a NLP tool that can automatically extract the 80 attributes presented in Table 1 and various safety outcomes with high accuracy (96% in *F*1 score). In this study, for illustration purposes (proof of concept), we apply our methodology on an attribute data set extracted from a pool of 5298 raw injury reports by the aforementioned NLP tool.

2.2. Modeling and managing safety in BIM

Among many characterizations, we refer to Building Information Modeling (BIM) as an information-rich design technology that can be used to generate a virtual model of an infrastructure. The strength of the BIM technology stems from its ability to augment the 3D representation of a facility with a plethora of information such as schedule,

Table 1
Attribute counts in our data set.

UPSTREAM	*Count*	Rebar	155	Screw	37
Cable tray	48	Scaffold	300	Slag	75
Cable	75	Soffit	12	Spark	9
Chipping	34	Spool	52	Slippery surface	142
Concrete liquid	58	Stairs	137	Small particle	401
Concrete	165	Steel sections	759	Adverse low temperatures	123
Conduit	56	Stripping	114	Unpowered tool	611
Confined workspace	129	Tank	85	Unstable support/surface	8
Congested workspace	13	Unpowered transporter	53	Wind	109
Crane	69	Valve	79	Wrench	110
Door	85	Welding	200	Lifting/pulling/manual handling	553
Dunnage	29	Wire	131	Light vehicle	133
Electricity	3	Working at height	268	Exiting/transitioning	132
Formwork	143	Working below elevated wksp/material	50	Sharp edge	47
Grinding	133	Drill	97	Splinter/sliver	41
Grout	18	**TRANSITIONAL**		Repetitive motion	66
Guardrail/handrail	91	Bolt	186	Working overhead	14
Heat source	111	Cleaning	119	**DOWNSTREAM**	
Heavy material/tool	79	Forklift	39	Improper body position	88
Heavy vehicle	143	Hammer	149	Improper procedure/inattention	57
Job trailer	24	Hand size pieces	172	Improper security of materials	87
Lumber	252	Hazardous substance	156	Improper security of tools	28
Machinery	189	Hose	95	No/improper PPE*	23
Manlift	66	Insect	105	Object on the floor	174
Stud	31	Ladder	163	Poor housekeeping	2
Object at height	86	Mud	35	Poor visibility	12
Piping	388	Nail	94	Uneven walking surface	59
Pontoon	15	Powered tool	239		

* Personal Protective Equipment.