Automation in Construction 94 (2018) 104–111

Contents lists available at ScienceDirect

Automation in Construction

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

Identification of latent legal knowledge in differing site condition (DSC) litigations

Tarek Mahfouz[a,*], Amr Kandil[b], Sukhrob Davlyatov[a]

[a] Department of Technology, College of Applied Sciences and Technology, Ball State University, AT 151 Applied Technology Building, Muncie, IN 47306, USA
[b] Division of Construction Engineering & Management, School of Civil Engineering, Purdue University, 550 Stadium Mall Dr., West Lafayette, IN 47907, USA

ARTICLE INFO

Keywords:
Differing Site Conditions
Legal knowledge
Knowledge management
Machine learning

ABSTRACT

Conflicts in construction projects have always been a major problem. Unless an alternate resolution mechanism is spelled out in the contract, these disputes are typically resolved in court, which might be time consuming and financially substantial. This paper represents a continuation in a research focused on creating robust methodologies for legal decision support within the construction industry. Consequently, this papers tackles the problem of automating the extraction of implicit knowledge about significant legal factors upon which verdicts of Differing Site Condition (DSC) litigations are based. To that end, the research methodology (1) utilized a set of 600 cases from the Federal Court of New York; (2) adopted 15 legal concepts that have been found to be statistically significant for DSC litigations; (3) implemented 4 weighing mechanism for data representation, namely Term Frequency, Logarithmic Term Frequency, Augmented Term Frequency, and Term Frequency Inverse Document Frequency; and (4) employed Machine Learning (ML) classifiers, namely Naïve Bayes, Decision Tree, and PART for the development of 12 prediction models. Among the finding of this study (1) ML classifiers present a suitable solution for the analyzed task; and (2) Naïve Bayes classifiers achieved the highest prediction accuracy of 88%.

1. Introduction

Two key issues that impact productivity in the construction industry, as emphasized by previous studies, are integration and knowledge management ([6, 8, 28], and [16]). These issues arise because (1) significant volume of information is included in text formats; (2) information analysis needed for situations assessments is time consuming and costly to obtain; (3) this construction data is not incorporated with other construction management systems and; (4) there is no clear relationship between construction data and related project elements [8]. Therefore, knowledge management in the construction industry is an essential yet challenging task. These challenges are also seen in legal domain, since legal documentation is typically stored in text files [4, 5]. Currently available electronic knowledge management systems for regulations, case histories, and laws are complicated, and are not user friendly which leads to issues for information seekers whose access to relevant datasets is, oftentimes, limited. The aforementioned exacerbates the complexity of taking decisions within the construction legal domain as it, now, may require a legal professional to seek the required information from relevant sources.

There is no shortage of claims and disputes in the construction industry. For instance, Ren et al. [26] states that: (1) 52% of all U.K. construction projects will involve a claim of some type; (2) At any point of time, the burden of construction claims on the industry could reach up to £1.2 billion; and (3) within the majority of project, about 83%, a minimum of one request of project deadline extension was granted. In the U.S. comparable statistics could be observed. In the last 30 years, the volume of claims in the construction industry has increased exponentially. Organizational, planning, and contractual problems are the primary reasons for the increase in construction claims [23]. In addition, many claims arise from (1) the projects' complexity levels; (2) strict pricing structures that forbid the mitigation of unforeseen costs; (3) and the contractual methods that dictate the extent of financial risks or exposure to each party [17]. Another study shows that the likelihood of construction claims is increased by the size and duration of the project, complexity of contractual agreements, lack of proper communication, financial issues, limited resources, poor design, labor disputes and force majeure events [13]. No matter what the reason is, construction claims have a significant negative impact on the construction process. If these claims are not resolved in a timely manner, they quickly escalate into very expensive and time consuming disputes that has a potential to damage the company's reputation [13, 17]. As the

* Corresponding author.
E-mail addresses: tmahfouz@bsu.edu (T. Mahfouz), akandil@purdue.edu (A. Kandil), sdavlyatov@bsu.edu (S. Davlyatov).

https://doi.org/10.1016/j.autcon.2018.06.011
Received 20 April 2017; Received in revised form 23 May 2018; Accepted 19 June 2018

construction industry plays a central role in developing infrastructures around the globe, these disputes, on larger scale, damage national economies. Specifically, the negative effect of these disputes is significant in larger and complex projects due to extraordinary expenditures associated with project delays, possible shutdowns, and cost of litigation itself [17]. According to [24] the annual burden of conflicts and disputes on the U.S. construction industry is estimated to in excess of $5 billion.

Unless, a construction contract includes a binding arbitration clause or an alternate resolution mechanism, construction disputes typically are resolved in the courts of law [14]. While, formal litigation process has several advantages, including the fact it is legally binding on both parties, it also has two major drawbacks, it could become an exceedingly time consuming and very expensive process. Depending on a jurisdiction, an intricate dispute may take two to six years before it is even tried. According to Treacy [27], within the period from 1984 to 1992, litigation cases that are stagnant within the system for at least three years have been doubled. In addition, the lengthy court proceedings render litigation process excessively costly. This argument is further strengthened because construction litigation process requires specialized individuals who possess advanced legal and construction experience. This skill set is not widely available in the industry [14]. Individuals with the mentioned skill set are limited, and, therefore, their billing rates are high-priced [11]. For instance, it has been reported that legal and expert fees in litigation proceedings has increased by 425% between 1979 and 1990. On the other hand, settlements and verdicts have only increased by 309% [22].

In an effort to find more effective and efficient ways to deal with construction disputes, researchers have endeavored to find mechanisms to mimic and model the reasoning process utilized in the judicial system by employing Artificial Intelligence (AI) technologies. Some of the previous developments include the use of (1) computer models that are based on if-then conditions and rules; (2) artificial neural networks (ANN); (3) similarity measures to previous cases through case based reasoning (CBR); and (4) hybrids between the aforementioned systems. While these developments are a massive success, there are shortcomings that need to be further addressed. One important shortcoming that needs to be addressed is that the previous systems were not based on detailed analyses of legal concept that govern the litigation process. These detailed analyses, arguably, are important because they largely determine the success rates of these litigation outcome prediction systems. Therefore, this paper aims at providing a solid methodology for legal decision making in the construction industry.

The focus of the paper is to develop automated extraction methods for legal factors commonly encountered in Differing Sites Condition (DSC) disputes employing machine learning (ML) modules. DSC disputes were chosen as the focus for this paper because of their frequency and weight in construction disputes. The current research endeavor, described within this paper, provides a continuation step in a research aiming at creating an inclusive Machine Learning (ML) and Statistical Modeling (SM) methodology for facilitating the prediction of DSC litigations outcomes. For more information about the ML and SM components, reference is made to [18, 21] respectively. To that end, this paper augments a previous research effort that assessed the suitability of Support Vector Machines (SVM) algorithms, [20], to automate the extraction of essential legal factors up on which judges base their decisions by creating (1) Naïve Bayes (NB), Rule Induction Classifiers including Decision Trees (DT), and the projective adaptive resonance theory (PART); and (2) comparing their prediction accuracy to the SVM algorithm and each other. The primary finding of the paper contributes to developing a coherent and integrated methodology for decision making process involving construction disputes employing statistical modeling and machine learning.

2. Background

Previous research in the AI domain shows attempts of researchers to employ logic in order to differentiate legal positions of parties and determine the root causes of their arguments. Initially, Diekmann and Kruppenbacher [12] attempted to develop a system the employs legal rules in the form of if-then conditions to analyze and evaluate the merit of construction claims. They employed the expert system (DSCAS) to analyze DSC claims. The model was based on the Federal Government Standard form General Conditions (2B-A, GP-4), while its outcome was the probability to pursue or drop a particular case based on the guidelines of the aforementioned conditions. In 1986, Cobb and Diekmann developed another system titled Claim Expert Knowledge System (CEKS) that was aimed at providing guidance to non-legal professionals [11]. CEKS was, essentially, the extension of DSCAS, because it was developed on expanded questions and answer sets. Similarly, the US Army Construction Engineering Research Laboratory (USA-CERL), in 1984 created a similar tool under the title of "Claim Guidance System (CGS-DSC)". The foundation of the system was the US Government Federal Contracts' DSC clause (FAR-52.236-2) [15].

However, these systems had low successful forecasting rates because they lacked inclusion of detailed analyses of mandatory rules that govern litigation proceedings [7]. In addition, the requirement of high computing power and maintenance were some of their drawbacks. Despite these shortcomings, these systems created a new school of thought that reinvigorated a debate and created a healthy challenge among researchers to develop a viable system. New approaches have been examined. Among these new approaches, Artificial Neural Networks (ANN) methodologies were created. Researchers employed ANN to model judicial thought process to predict an outcome of a dispute ([3, 9], and [10]). These systems employed a variety of inputs including project changes, contract type, conditions, involvement of parties etc. The output of the system was the court's decision. Calculation of these systems ranged from 80% and 84% achieved by Chau [9] and Chau [10], and 67% achieved by Arditi et al. [3].

In addition, a new AI technique titled Case Based Reasoning (CBR) was developed. In 1999, Arditi and Tokdemir developed a CBR prototype to predict outcomes of construction cases, which attained an accuracy rate of 83% [2]. More recent advancements in AI research enabled for higher precision rates. In 2010, Arditi and Pulket achieved a prediction accuracy of 91.15% through the development of an Integrated artificial intelligence Prediction Model (IPM), which utilizes data consolidation, attribute selection methodologies, as well as multiple hybrid classification systems [1]. A higher prediction rate of 96% was attained by the authors through the use of Support Vector Machines (SVM) algorithms for the prediction of Differing Site Conditions litigations outcomes within the New York Federal Court [18].

However, with the great achievements of the aforementioned research efforts, their predictions are contingent on the accuracy of the factors used to derive such predictions. Thus, automating the process of identifying and extracting these factors from textual documents based on the analysis employed by judges is essential to advance these models and provide a robust comprehensive methodology for the construction industry. Thus the purpose of this paper is (1) investigate the suitability of automating the extraction of pertinent legal factors from textual documents through creating Naïve Bayes (NB), Rule Induction Classifiers including Decision Trees (DT), and the projective adaptive resonance theory (PART); (2) comparing their prediction accuracy to previously developed SVM models [20]; (3) identifying the best model with the highest accuracy rate; and (4) testing and validating the best model.

3. Methodology

As mentioned earlier, this paper provides an expansion in a lines of research that focuses on providing robust and comprehensive method