

Natural Language Processing for Information and Project Management



Giuseppe Martino Di Giuda, Mirko Locatelli, Marco Schievano,
Laura Pellegrini, Giulia Pattini, Paolo Ettore Giana and Elena Seghezzi

Abstract The purpose of the paper is to investigate the state of the art of textual translation theories, methods, and tools into formal and numerical requirements to support information modelling and project management process. Natural Language Processing helps translate text requirements into numerical terms that are necessary for the application and success of information modelling and management in a data-driven process.

Keywords Project and information management · Natural language processing · Data-driven process · Artificial Intelligence

1 Introduction

The purpose of the article is to investigate the state of the art about theories, methods, and tools for translating documents into numerical requirements and possible applications in the field of information modeling and project management. Information modeling requires a precise and comprehensive definition of the initial requirements; the computational definition of the initial requirements is fundamental for the good application of information modeling and management. Natural language processing (NLP) can be applied to translate text into numerical data. In order to understand the proposed methodology and its possible applications, it is necessary to explain basic theories and the most recent developments and applications of NLP. We present case studies of applications in various fields focusing on possible applications in the AECO sector, with a focus on possible data-driven management of the construction process. A discussion on possible applications and further developments is also presented.

G. M. Di Giuda (✉) · M. Locatelli · M. Schievano · L. Pellegrini · G. Pattini · P. E. Giana · E. Seghezzi

Architecture, Built Environment and Construction Engineering—ABC Department, Politecnico di Milano, Milan, Italy

e-mail: giuseppe.digiuda@polimi.it

© The Author(s) 2020

B. Daniotti et al. (eds.), *Digital Transformation of the Design, Construction and Management Processes of the Built Environment*, Research for Development, https://doi.org/10.1007/978-3-030-33570-0_9

2 Natural Language Processing (NLP): Rule Based, Statistical and Deep NLP

Natural Language Processing (NLP) is an interdisciplinary field involving humanistic, statistical–mathematical, and computer skills. The aim of NLP is to process languages using computers. The human language can be defined as natural because it is ambiguous and changeable. On the contrary, machine language is defined as formal because it is unambiguous and internationally recognized. The NLP must deal optimally with the ambiguity, imprecision, and lack of data inherent in natural language. NLP tasks may take two main approaches: a machine learning (ML)-based approach (i) or a rule-based/statistical approach (ii). An ML-based approach uses ML algorithms for text processing (Pradhan et al. 2004), whereas a rule-based approach uses manually coded rules (Soysal et al. 2010). NLP approaches could be either shallow or deep. Shallow NLP conducts partial analysis of a sentence or extracts partial, specific information from a sentence. Deep NLP aims at full-sentence analysis toward capturing the entire meaning of a sentence (Zouaq 2011).

A table comparing the different approaches of the NLP has been created in order to identify the most suitable one for the numerical translation of the requirements (Table 1).

As shown below, the statistical approach to the NLP seems to have more advantages than traditional/manual rule based. However, among the various NLP techniques, those based on Artificial Neural Networks seem to have more chance of success.

ML-based methods efficiently manage the sparsity and non-structuring of learning data. Among the techniques of ML for NLP, those based on ANN ensure computational processing of text documents better, because they can respect the complexity, articulation, and multidimensionality of human natural language. By translating text into machine language, text information can be managed and used with methods, techniques, and tools typical of project and information management. In summary, NLP, based on ANN, transforms text documents into structured information resources (Callison-Bourne and Osborne 2003).

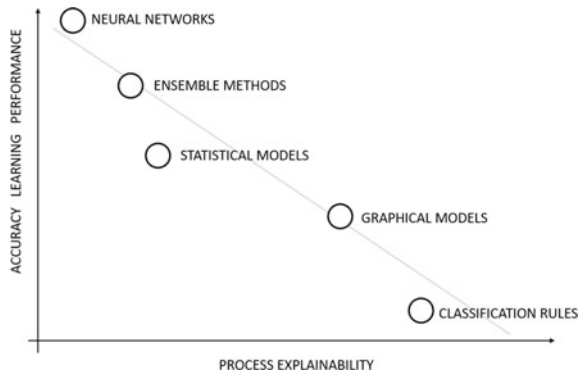
Despite the numerous advantages of the ANN compared to the rule based and statistics NLP, it offers few explanations on the relations found between the data and the connection among neurons with the output data. The phenomenon is called black-box effect which makes difficult to explain what is learned from the net (Paliwal and Kumar 2011; Waziri et al. 2017).

The latest efforts in the field of AI aim at overcoming the black box effect and the production of explainable AI model called the XAI model (Gunning 2017) (Fig. 1).

Table 1 Advantages and disadvantages of the various NLP approaches

NLP approach	Advantages	Disadvantages
Rule based	It produces outputs in relation to the data entered	Long time to implement and test a prototype
		It produces outputs in relation to the data entered
		Manual system creation, time-consuming, and expensive creation process
Statistical based	Rapid prototyping and testing	Robust system (Junqua and Van Noord 2001): it will always produce outputs, regardless of the inputs entered
	Robust system (Junqua and Van Noord 2001): it will always produce outputs, regardless of the inputs entered	
	Automated system creation, cheap creation process	
Artificial Neural Network based	Rapid prototyping and testing	ANNs operate as a black box (Jain and Pathak 2014)
	Robust system: ANNs can manage noisy and ambiguous data, prone to error or incomplete (Pradhan et al. 2004; Sivanandam and Deepa 2006; Rumelhart 1986)	
	Structure similar to the human brain, ANNs are very effective for solving a complex problem	
	ANNs have the ability to overcome the lack of data, they “Fill in the blank”. This makes it particularly suitable for use in the NLP (Waziri et al. 2017)	

Fig. 1 Explainability and performance qualitative graph of AI learning method (Gunning 2017)



3 NLP Application for Project Management in AECO Sector

The increasing complexity and size of projects in the AEC sector make it difficult to identify and verify initial requirements expressed in natural language. As a result, there are numerous errors and shortcomings in the definition of requirements at the early stages of concept and design, with inefficiencies in terms of time, cost and quality (Frenette and Kyriakidis 2016). The use of NLP can help the project manager and the client to express project requirements in alphanumeric and quantifiable terms avoiding misunderstanding and increasing the project's chances of success. The use of NLP in the early stages of defining project requirements can be therefore considered a risk mitigation technique.

ANN approach to NLP allows to define requirements and predictions in order to monitor the progress of the project. This, in turn, leads to an optimization of the definition of requirements and the assessment of the project progress.

Alphanumeric translation of text documents into formal and structured data is a prerequisite for information modeling. The initial phase of defining requirements is fundamental for the good application and effectiveness of the information modeling method. The NLP for requirements definition can be integrated with the information modeling and management of the preliminary phase of the design and construction process (Zhang and El-Gohary 2015).

NLP approach can reduce time and cost in defining requirements by avoiding errors, loss of information and ambiguities in defining project objectives (Zhang and El-Gohary 2015).

The NLP approach can support Automated Compliance Checking in a BIM-based process (Zhang and El-Gohary 2015). Construction projects must comply with a variety of standards code. The manual conformity control process is time-consuming, costly, and error-prone (Han et al. 1998; Nguyen 2005). Automated Compliance Checking (ACC) and NLP should reduce Code Checking time, costs, and errors (Salama and El-Gohary 2013; Tan et al. 2010).

Some authors (Zhang and El-Gohary 2015) point out that, in addition, the proposed method has other potential benefits:

- allowing potential non-compliance cases to be identified in advance, which could save significant time and cost due to changes and/or rework (Ding et al. 2006),
- promoting the adoption of Building Information Modelling (BIM) and increase the cumulative benefits of adopting BIM as BIM would allow ACC (Pocas Martins and Abrantes 2010),
- enabling more efficient integration of stakeholder inputs into the design and exploration of what-if design scenarios. As a result, experimenting different design options and checking for compliance would be more efficient in terms of time (Frenette and Kyriakidis 2016),
- reducing violations of regulations, due to easier and more frequent Compliance Checking (CC) (Zhong et al. 2012).

The survey carried out shows that NLP can support the automatic identification of the limits, according to the regulations and support Automated Code Checking. This method applied to the BIM model can lead to a substantial reduction in time and errors (Zhang and El-Gohary 2015).

4 An Application: NLP for Risk Management

This paragraph proposes several applications of NLP in a specific construction related field. Risk management was chosen, as it reflects several declinations of NLP use.

Referring to safety risk management, NLP is applied to Case Base Reasoning (CBR). CBR is an important approach in safety risk management of construction projects. It emphasizes that previous knowledge and experience of accidents and risks are extremely valuable and could help to avoid similar risks in new situations. CBR allows the creation and increasing size of Databases on construction accidents. These documents are written in natural language; as a consequence, retrieving information quickly and accurately from the database is still the main challenge. In order to improve the efficiency and performance of data recovery, a Natural Language Processing approach is proposed (Zou et al. 2017).

NLP gives the advantage to ensure an easier recovery of CBR knowledge and cases.

As regard to construction risk management, NLP approach is used to analyze bid document.

The bidding process takes place in the early stages of a construction project. Bidders should fully understand the uncertainties of the project before making decisions. The risk of uncertainty of construction projects, in turn, is determined by the content of the bid document. If the information provided in the bid notice is not accurate and clear, the uncertainty of the projects increases (Lee and Yi 2017).

NLP approach is proposed to predict risks in the bidding process of construction projects, by analyzing the uncertainty of the bidding document and using it as a factor to predict the bidding risk of a project. The model for forecasting tender risk was conducted using the pre-bid clarification information. Text mining was carried out on pre-bid documents, which are in an unstructured text data format, and the results were used as the main influencing factors for the risk forecasting models (Lee and Yi 2017).

In the field of safety and pre-bid risk mitigation, NLP is a useful method and tool to support the project manager in project risk mitigation.

The last application of NLP refers to procurement risk management.

As the size and complexity of construction projects significantly increased, the number of disputes among the parties involved during construction work is constantly increasing. To avoid such disputes, participants need to be sure of their contractual positions and rights. Most international construction projects require contract management teams to examine all possible contract risks during tendering periods. However, it is very difficult to review a large number of contracts in a short period of

time. NLP approach is used to propose a model of automatic extraction of poisonous clauses (Lee et al. 2019).

The NLP approach proposed is, therefore, effective in automatically reading and extracting poisonous clauses from construction contracts. The algorithm, suitably modified, can be used to read and extract data from documents other than construction contracts.

5 Conclusions and Further Developments: NLP for Requirements Engineering

Given the importance of the correct and complete definition of the initial project requirements for the successful application of project and information management, the NLP approach to support the translation of the requirements into numerical and alphanumeric terms could help reducing the gap between expected and actual quality. The presented case studies show that, methods, techniques, and application cases of NLP have already been tested. However, no case of systemic application to construction project has been found since the early stages of planning. The use of NLP for the definition of initial requirements would have a greater impact if applied at an early stage. The NLP approach could be used for the definition of the initial requirements of a public client helping the public actor to define requirements on a numerical and alphanumeric basis and not on a simple text basis. Through the alphanumeric requirements translation, the public client would be the main actor in a Data-driven construction process, increasing its ability to understand, manage and direct the outcome of the design. The numerical translation of the requirements would make the entire process computable, and digitally manageable, which would see the client as able to contribute directly to the design process (Ciribini 2016). NLP supports requirements engineering by transforming the classic qualitative demand based on text data into a computational demand based on formal and structured data. During a data-driven process, the monitoring of the objectives to be achieved can be more effective and immediate, reducing the risk of overcoming time and costs and not reaching the expected quality level.

Possible further development of requirements translated into a computational form, through NLP, is the possibility of being readable and digitally managed. In this way, the design and construction process can be managed digitally from the earliest stages. Text translated into informal language data using NLP approach can be used to train an Artificial Intelligence. The A.I. needs a huge quantity of data to be trained. The amount of data needed for training A.I. can be an application limit. The AI, if properly instructed, could support the evaluation phase of the project to verify the degree of compliance of the project with the requests coming from the translation of text document. The system structured in this way would be configured as a decision support system for assessing the progress of the project. The monitoring

of the project progress would be automatically managed by the AI to avoid errors and costs in economic and temporal terms.

References

- Callison-Bourne C, Osborne M (2003) Statistical natural language processing. In: Farghaly A (ed) A handbook for language engineers. CSLI Publications
- Ciribini ALC (2016) Information modelling management, BIM e digitalizzazione dell'ambiente costruito. Grafill S.r.l, Palermo
- Ding L, Drogemuller R, Rosenman M, Marchant D, Gero J (2006) Automating code checking for building designs Design check. Clients driving innovation: moving ideas into practice. CRC for Construction Innovation, Brisbane, Australia, pp 1–16
- Frenette M, Kyriakidis J (2016) Rethink requirements—The Natural language processing approach. Project management.com, 12 July. <https://www.projectmanagement.com/webinars/334316/Rethink-Requirements—The-Natural-Language-Processing-Approach>
- Gunning D (2017) Explainable artificial intelligence (xai). Defense Advanced Research Projects Agency (DARPA)
- Han CS, Kunz JC, Law KH (1998) Client/server framework for online building code checking. J Comput Civ Eng 12(4):181–194. 10.1061/(ASCE)0887-3801(1998)
- Jain M, Pathak KK (2014) Applications of artificial neural networks in construction engineering and management: a review. Int J Eng Technol Manag Appl Sci 2(3):134–142
- Junqua J-C, Van Noord G (2001) Robustness in language and speech technology, vol. 17 of Text, speech and language technology. Kluwer Academic Publishers
- Lee JH, Yi JS (2017) Predicting project's uncertainty risk in the bidding process by integrating unstructured text data and structured numerical data using text mining. Appl Sci 7(11):1141. <https://doi.org/10.3390/app7111141>
- Lee JH, Yi JS, Son JW (2019) Development of automatic-extraction model of poisonous clauses in international construction contracts using rule-based NLP. J Comput Civ Eng 33(3):04019003. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000807](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000807)
- Nguyen T (2005) Integrating building code compliance checking into a 3D CAD system. In: Proceedings of international conference on computing in civil engineering, ASCE, Reston, VA, pp 1–12
- Paliwal M, Kumar UA (2011) Assessing the contribution of variables in feed forward neural network. Appl Soft Comput 11(4):3690–3696
- Pocas Martins JP, Abrantes V (2010) Automated code-checking as a driver of BIM adoption. J Hous Sci 34(4):286–294
- Pradhan S, Ward W, Hacioglu K, Martin JH, Jurafsky D (2004) Shallow semantic parsing using support vector machines. In: Proceedings of the NAACL-HLT, association for computational linguistics, East Stroudsburg, PA, pp 233–240
- Rumelhart DE (1986) Learning internal representations by error propagation. Parallel Distrib Process 1:318–362
- Salama D, El-Gohary N (2013) Semantic text classification for supporting automated compliance checking in construction. J Comput Civ Eng: 04014106. [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000301](https://doi.org/10.1061/(asce)cp.1943-5487.0000301)
- Sivanandam SN, Deepa SN (2006) Introduction to neural networks using Matlab 6.0. Tata McGraw-Hill Education, Columbus, United States
- Soysal E, Cicekli I, Baykal N (2010) Design and evaluation of an ontology based information extraction system for radiological reports. Comput Biol Med 40(11–12):900–911
- Tan X, Hammad A, Fazio P (2010) Automated code compliance checking for building envelope design. J Comput Civ Eng 24(2):203–211. 10.1061/(ASCE)0887-3801(2010)

- Waziri BS, Bala K, Bustani SA (2017) Artificial neural networks in construction engineering and management. *Int J Arch Eng Constr* 6(1). <https://doi.org/10.7492/IJAEC.2017.006>
- Zhang J, El-Gohary NM (2015) Automated information transformation for automated regulatory compliance checking in construction. *J Comput Civ Eng* 29(4):B4015001. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000427](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000427)
- Zhong BT, Ding LY, Luo HB, Zhou Y, Hu YZ, Hu HM (2012) Ontology-based semantic modeling of regulation constraint for automated construction quality compliance checking. *Autom Constr* 28:58–70
- Zou Y, Kiviniemi A, Jones SW (2017) Retrieving similar cases for construction project risk management using natural language processing techniques. *Autom Constr* 80(September 2016): 66–76. <https://doi.org/10.1016/j.autcon.2017.04.003>
- Zouaq A (2011) An overview of shallow and deep natural language processing for ontology learning. In: *Ontology learning and knowledge discovery using the web: challenges and recent advances*. IGI Global, Hershey, PA, pp 16–38

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

