Text Mining Post Project Reviews to Improve the Construction Project Supply Chain Design

A. K. Choudhary¹, J. A. Harding¹*+, P. M. Carrillo², P. Oluikpe² and N. Rahman¹
¹Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University, Loughborough, Leicestershire, United Kingdom
²Department of Civil and Building Engineering, Loughborough University, Loughborough, Leicestershire, United Kingdom
*Corresponding Author: J.A.Harding@lboro.ac.uk

Abstract - Post Project Reviews (PPR) capture good and bad practices, identify problems, waste, risks, missed opportunities, communication lag, financial issues, partner relationships etc., of a construction project supply chain (CPSC). They are huge sources of information, knowledge and experience from project managers, clients, suppliers and contractors, related to issues from every stage of the construction project. If these reports were analysed collectively, they may expose important detail, perhaps repeated across a number of projects. However, because most companies do not have the resources to thoroughly examine these PPRs, either individually or collectively, important insights are missed thereby leading to missed opportunities to learn from previous projects. This research shows that the hidden knowledge and experiences could be captured using knowledge discovery and text mining approaches to uncover patterns, associations, and trends in PPRs. The results might then be used to address specific problem areas, enhance processes and improve the design and planning for new construction projects.

Keywords: PPRs, Construction Project, Supply Chain, Knowledge discovery and Text Mining

1 Introduction

The construction industry forms one of the most diverse and unstable sectors within the UK economy. It faces widely fluctuating demand cycles, project specific product demands, uncertain production conditions and has to combine a diverse range of specialist skills within geographically dispersed short-term project environments[1]. A construction project supply chain (CPSC) may contain hundreds of firms, contractors, subcontractors, material and equipment suppliers, engineering and design teams, and consulting firms[2]. Collaborations between the various entities of the CPSC are temporary and may vary from project to project. The lifecycle of a construction supply chain is limited to a particular project only. Post Project Reviews (PPR) of construction projects are one of the most important and common approaches for the capture of knowledge and lessons learned from the operation of a CPSC. They provide opportunities for the project team members to share, discuss and even explain their experiences through face-to-face, facilitated interactions before a project is closed and the team is dissolved. PPRs therefore allow multi-disciplinary teams to critique a project to determine both positive and negative aspects, potentially capturing tacit knowledge as learning points to improve the planning, execution and design of new construction projects [3-4]. Debates between the project team members during PPRs may lead to greater innovation and better ideas than can be achieved from any individual. Orange et al. [5] and Kamara et al. [6] also identified that PPRs have a huge potential for much more thorough exploitation. If information and knowledge from PPRs can be extracted and analyzed effectively, good and bad practices might be identified so that lessons are learnt from past projects and knowledge reused or exploited to improve the quality and levels of success in redesigning the CPSC.

Knowledge Discovery in Text (KDT) and Text Mining (TM) [12] are very recent and increasingly interesting areas of research in computer science. KDT and TM are mostly automated techniques that aim to discover high level information from huge amounts of textual data and present it in a useful form to the potential user, who might be an analyst, decision maker or project manager etc. KDT or TM techniques might therefore be applied on PPR documents from multiple projects to potentially identify information relating to bad, good or even best practices. The benefits of mining PPR texts lie in the possibilities of discovering patterns, associations and linkages of processes, activities and terms occurring in the reports. The organization may then adjust its activities to reflect what is learned from the KDT and TM with the aim of improving processes, optimizing profit and improving client retention. In addition, whenever a new construction project is initiated, it would be very beneficial if lessons learned from previous projects could be quickly identified to reduce the chances of errors being repeated and increase the potential for savings in cost and time.

This paper examines a method for applying TM on PPRs as follows: Section II discusses how PPRs capture the knowledge and experiences of CPSCs; Section III briefly discusses Knowledge Discovery in Text and Text Mining...
technology. Ontology development for PPR to facilitate the text mining is discussed in Section IV and Section V examines the application of text mining for discovering new knowledge from PPRs. Section VI illustrates examples from two UK construction companies. Section VII concludes the paper with a direction for future research.

2 Post Project Review of Construction Project Supply Chain

2.1 Construction Project Supply Chain

CPSC involves all processes, activities, tasks and information flows (both upstream and downstream) within various networks of organizations involved in the delivery of quality construction projects or services. A general CPSC may contain several firms, contractors, subcontractors, material and equipment suppliers, engineering and design firms or teams, consulting firms or teams, etc. It remains highly fragmented and involves many small and medium size suppliers and sub contractors [2,7].

2.2 PPRs of Construction Project

PPR of a construction project is a process through which a construction company looks at the various stages of its supply chain retrospectively with a view to learning from activities carried out, to avoid mistakes in the future and also to learn from successes and failures[9]. An example of the linkages between a construction supply chain and PPR is shown in figure 1. Lessons learnt may be used to redesign and benefit the future construction projects. Benefits gained by organizations from conducting PPRs have been highlighted in Tan et al. [4] and Carrillo [3] and include:

- **Facilitating collective learning:** PPRs provide an opportunity for people involved in the various stages of a construction project to come together and examine what went right or wrong during the project or during a particular stage of the project through knowledge sharing, exchange of ideas, brainstorming and contributions.

- **Provide utilisable knowledge:** PPRs embody the knowledge of the project and as much of the knowledge comes from the project contributors, it is often tacit knowledge and therefore can be difficult to reuse.

- **Benefit client organizations:** Review processes often aim to provide greater insight into how an organization functions in managing its assets. This helps the project organization develop its processes and also manages its assets better.

- **Better project phase management:** Reviewing each phase of a project provides opportunities for better project management at the phase level, rather than carrying out a single review at the end. Hence, mistakes might be corrected earlier at the phase level perhaps benefiting the remaining project phases.

- **Prevent knowledge loss:** When a project team disbands, most staff are reassigned to other projects and they carry with them valuable knowledge about the project. It is therefore important that a PPR process captures this knowledge about the project and makes it explicit for others to utilize.

Newell [10] pointed out that the PPRs of construction projects are a huge silo of information which rarely gets analyzed critically to reveal patterns of information that could help

![](image-url)

Fig. 1. Generic view of Construction Supply Chain and linkages with PPR.

The upstream responsibilities of the CPSC consist of the activities and tasks leading to the preparation of the production on site involving the construction client and design team. The downstream activities and tasks involve the construction suppliers, subcontractors and specialist contractors. The CPSC therefore needs a high level of coordination among various stakeholders, who may have conflicting interests. This paper only considers the physical view and therefore does not take into account the management philosophy of supply chains. Construction supply chain management is not the focus of this paper and indeed very little work has been carried out in the area of CPSC management [8].
decision making in the project process. Due to this inability to convert the review contents into useful knowledge, project teams abandon the report on the shelf and move on, thus ignoring knowledge which might be useful and even critical in designing a new CPSC.

The major limitations that exist in exploitation and dissemination of PPRs can be summarized as follows:

1. Knowledge and learning from previous PPRs are not routinely transferred to future or ongoing projects and therefore learnt lessons are not properly exploited or reused.

2. Collections of PPRs are not commonly analyzed or explored together to find the hidden knowledge, recurring problems or patterns of behavior that may exist within them.

3. PPRs may identify good and bad practices but existing PPR processes do not detail how these should be disseminated in order to improve new CPSC design.

4. PPRs commonly generate large quantities of documentation which may be stored on a company network. However research is still required to determine how to effectively disseminate analyzed PPRs throughout an organization.

Drawing from the above, a major challenge for construction organizations is how to extract knowledge from PPRs and disseminate this appropriately across the organization to ensure optimum improvements in the design of CPSCs. The application of KDT and TM methods on PPRs should provide opportunities for better exploitation of the potential benefits of these PPRs by extracting useful information to identify and replace poor practices, improve process performance, avoid reinventing solutions, re-use lessons learned on previous projects etc. The next section briefly discusses KDT and TM methods.

3 Knowledge Discovery in Text and Text Mining

KDT and TM involve techniques like information retrieval, text analysis, information extraction, clustering categorization, visualization, database technology, machine learning, natural language processing, data mining [11] and knowledge management [12]. Following the definition of KDD by Fayyad et al. [13], Karanikous and Theoudoulidis [12] defined KDT as “the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in unstructured data”. On the other hand, Text Mining (TM) is a step in the KDT process consisting of particular data mining and natural language processing algorithms that under certain computational efficiency and limitations produce a particular enumeration of patterns over a set of unstructured textual data. Figure 2 shows a generic view of a text mining process from unstructured or semi structured textual data.

![Knowledge discovery in text and text mining process](image)

The following major strengths and benefits of text mining [14] could be used to address the identified limitations and research gaps of PPR processes previously mentioned in this paper.

- TM can be used to extract relevant information of different types from one or more documents.
- TM can be used to gain insight about trends, relationships between people/places/organizations etc., by automatically aggregating and comparing information extracted from the documents of a certain type.
- TM can be used to classify and organize documents according to their content. PPR documents might therefore be classified using TM so that at the start of a new project, the most appropriate documents are automatically pre-selected into groups on a specific topic, analysed for appropriate knowledge and lessons learnt and then assigned to an appropriate person from the new project team.
- TM tools and techniques such as link terms and clustering can be used to retrieve documents based on various sorts of information about the document content.

To date, very little work has been published in the area of text mining applications [15-16].

4 Ontology Development for Text Mining

It is necessary to determine an ontology of important and common terms within the PPR reports in order to facilitate the processes of knowledge search and knowledge discovery. In practical terms, to satisfy the requirements of simple exploratory TM experiments, it would be sufficient to analyse
example PPR reports, find the common terms (which indicate the types of knowledge that are likely to exist in the reports) and then combine and refine these outputs with information provided by our collaborating companies related to key areas of interest. The results from this analysis and refinement could then be used to direct the text mining experiments to identify any useful knowledge that may lie hidden within the reports. However, we hope to be able to move beyond this type of explorative text mining experiments and determine a more specific approach for knowledge discovery within PPR reports, so that particular types of knowledge can be targeted for knowledge discovery. With this in mind, we have extrapolated and developed an ontology based on the combination of key terms identified in the example reports provided by two companies.

A set of hierarchies has been developed to capture the various issues identified within PPRs of CPSC. Each hierarchy has a single root which indicates the main topic areas shown in Figure 3. The Top Level hierarchy shows that the Learning or Outcome from a project will relate to one or more of the following topics, Finance, Quality, Communication, Building, Health and Safety, Labour, Environment, Time, Security.

![Top Level hierarchy relating to construction PPRs.](image)

Materials, Plant or Project Stages of a construction supply chain. Each of these areas is then considered in further detail in separate hierarchies, using key words from items of learning identified in the example PPR reports. For example, figure 4 shows the hierarchy for finance related key words.

![Hierarchy relating to finance from top level hierarchy](image)

In order to focus the text mining on particular topics of interest it is likely that two or more hierarchies will be utilised simultaneously. For example, to identify knowledge about delays on particular parts of buildings, both the “Time” hierarchy and the “Buildings” hierarchy would be used.

5 Text Mining PPR

In order to address the recurring problems of PPRs identified in section 2, there is a need for better exploitation and reuse of the information and knowledge stored within the PPR documents. Text mining has huge potential for addressing the identified problems of PPRs and in particular it can be used as an effective tool to support the following six tasks in the context of PPRs:

- Extracting common patterns of good or bad practices
- Finding correlation or linkages between commonly occurring terms, for example profit vs. regional targets.
- Clustering or grouping of reports based on pre-determined criteria.
- Summarization and generation of statistics to give a brief overview of groups of reports.
- Search and retrieve a particular context oriented PPR report., and
- Dissemination of analyzed reports across the organization using the world wide web.

To achieve useful results from TM, it is necessary to combine both domain expertise and text mining expertise. Figure 5 shows the linkages between these two types of expertise and they will now each be discussed in detail.

![Linkage between Domain expertise and Text mining expertise](image)

**Domain Expertise**: Table 1 shows the key topics and typical information found in a PPR report of a CPSC. This information was gathered through reviewing the PPR
processes of two construction companies. The left hand column relates to the high level ontology developed in Figure 3.

Table 1: An Example of Content of Domain Expertise

<table>
<thead>
<tr>
<th>General outcome</th>
<th>Financial, purchase, Contract period variation</th>
<th>How did the job go? ( \text{Profit or Loss} )? Extension of project period?</th>
<th>Were resources effectively allocated? Quality of service, environmental issues etc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Any delay? Reason for delay such as weather, changes or unforeseen events. What is required to finish the work on time? etc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality</td>
<td>Any defect during project? Redesigning or repair, any specific problem, leaking, faults, errors, mistakes, damage, snags or hiccups? How we achieved high quality? Was quality combined with speed?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication</td>
<td>Was there clear communication between management, client and the project team? How was the interaction with design team? What did we do well? What went wrong? What major changes arose from meeting with clients? Was the customer’s feedback, positive or negative?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Building</td>
<td>Issues involved with building such as drains, floors, cladding, beams, frames, ceiling, glazing, drains, partitions, walls, lifts etc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Security issues</td>
<td>Any material theft or loss and measures to prevent them. Which security methods (CCTV, movement sensors or security guards) worked well?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Material issue/plant issue</td>
<td>How Issues such as waste, delivery of materials, disposal and recycling are dealt with. Any problems that occurred during purchasing or procurement stage?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labour</td>
<td>Team/sub contractor/contractor/consultant</td>
<td>Any problems or concerns related to designer, supplier, engineers, consultants, surveyors, architect and builder, which effected the CPSC. Did designer and contractor work well together?</td>
<td></td>
</tr>
<tr>
<td>Project stages:</td>
<td>pre contract, lease, negotiation, agreement, site survey, estimation, design, tender, contract,</td>
<td>What sort of practices were adopted during the various stages of the project? What were good or bad practices used? Were the documents clearly communicated?</td>
<td></td>
</tr>
<tr>
<td>Project stages: Planning</td>
<td>Any comments for or from the planners? Was the sequence shown on programme correct? Was the duration reasonable?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project stage: Procurement</td>
<td>Was there any problem in getting subcontractors on time?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health and Safety: accidents, incidents, hazards</td>
<td>Are there any reported accidents or incidents? What are the causes? How did we perform over all? Any health and safety lessons to be learnt from this project?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Changes: plan, schedule, contract, order deadline, design, personnel, client or specification change etc.</td>
<td>What changes were made in the project and why? How could they be avoided in future projects?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mistakes/Errors</td>
<td>Were there any notable mistakes made on the site? What was the cause? How can these be corrected?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waste/environmental issues</td>
<td>Were any particular operations wasteful on time or material? Could the operation be improved by changing the material or method?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Innovation</td>
<td>Did any innovative or interesting ideas emerge during the whole project?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Text Mining Expertise: There are several commercial products available for text mining. Table 2 lists the various functionalities performed by different software packages as follows.

Table 2: Comparative Study of Text Mining Softwares

<table>
<thead>
<tr>
<th>Product</th>
<th>FE</th>
<th>TBN</th>
<th>SR</th>
<th>TC</th>
<th>Clus.</th>
<th>Sum.</th>
<th>Asc</th>
<th>InV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Smart discovery</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Text miner 1.0</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Autonomy</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 SAS Text miner</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Clear forest</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Polyanalyst 5.0/Text Analytics2.3</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Intelligent Miner</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Retrieval ware</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Examples of the application of several of these TM operations on PPR documentation will now be given to further illustrate the potential benefits which may be obtained and facilitate better knowledge reuse and exploitation from PPR processes. In the next section, an illustrative example has been presented showing how Polyanalyst [17] could be used to address the six exploitation and reuse requirements of PPRs.

6 A Case Study of UK Construction Projects

This example will consider PPR documentation collected during the last three years from 2 construction companies. Although the style of the reports from the different companies varies, each report contains textual narration and a description of the project, with the review divided under different headings and textual information providing the lessons learned during the operation of the CPSC.

6.1 KDT and Text Mining Process on PPRs

6.1.1 Transformation, Loading and Pre-processing

25 PPRs with page lengths varying from 15-30 pages were selected and imported to an Excel file. The file was then loaded into the Polyanalyst software system. Pre-processing is mainly done to reduce information overload and generate metadata. Textual data pre-processing steps include removal of “unwanted and non informative” words and stemming.

6.1.2 Text-Mining of PPRs

After pre-processing, the PPRs are ready for TM, which involves using various tools and techniques to extract patterns, trends and useful information. The following subsections will discuss the six major TM tasks (listed in section V) to address current problems commonly identified in PPRs.

Summary statistics and generation of statistics for CPSC:

Generally the summary statistics and summarization tools of TM are used for this purpose. They are discussed as follows:

Summary Statistics: Basic statistics can be generated for the PPR text at various stages of the TM to compare its attributes, key words, or generated rules. In the present example, these statistics help in finding the frequencies of frequently used words in the PPR reports. Thus, summary statistics are useful in identifying which reports are most important in the context of a particular issue of the supply chain.

Summary: Summarization techniques condense the PPRs to a fraction of their original size yet still retain the significant content of the PPRs in the summary. Summarization techniques determine the semantic weight of sentences written in the PPRs and only those sentences whose semantic weight is higher than the chosen threshold are kept. The size of the summary can be modified by changing the
Extracting common patterns of good or bad practices across supply chain: For this purpose, Text Analysis (TA) provides the morphological and semantic analysis of unstructured textual PPR reports. TA extracts and counts the most important words and word combinations from the textual PPR reports, and stores terms-rules for tokenizing the PPR records with the patterns of the encountered terms. The terms or a combination of terms are generated as rules which record the number of times each term or combination of terms exist and where the occurrences are. These rules can be applied on the PPRs to find patterns of events or activities, issues, causes or achievements in various areas of CPSC such as planning, change, estimation, errors or mistakes, quality, health and safety, defects and many more. Identifying problems, issues and possibly their causes in this way may help managers to avoid them while designing the future CPSC.

Finding correlation or linkages between commonly occurring terms: The commonly used technique for this purpose is Link Term (LT) analysis. It visually represents the complex patterns of relations between key words in the textual report. This mechanism provides the quickest way to understand the most prominent semantic characteristics of the explored textual reports. This analysis also enables relevant subsets of the original PPR data to be collected for further exploration based on the particular topics or terms of interest across the CPSC. For example, it will be able to show how modifications or changes in the schedule delayed the delivery of the project.

Clustering or grouping of reports based on defined criteria: Text categorization (TC) can be used to automatically split up the PPR reports into more homogeneous clusters based on keywords found in the PPR report. This tool therefore enables important clusters of reports (or pieces of text) to be identified as expressing similar ideas, issues, or concepts. TC is an important tool of TM, which automatically builds a hierarchical tree-like taxonomy of topics and subtopics extracted from unstructured textual reports.

Search and retrieve a particular context of CPSC from PPR reports: In situations where the new CPSC manager is looking for knowledge on a particular topic or in a particular context TM provides useful facilities to aid the semantic search. It is very similar to natural language queries, where queries are made using natural language by typing a question in conversational English. The result pane will display all the related answers to the query. Results are displayed as a tree-like structure based on the question. This sub-tree of concepts that are related to the query may help the user in simulating a better answer to the asked query.

Dissemination of analyzed reports across the members of a supply chain: All the reports generated by the software system could be exported as “.htm” files, which can be distributed to various members of the supply chain via the world wide web (providing appropriate id, password and access rights are set up for these individuals). TM therefore facilitates the effective dissemination of learning from PPRs in two ways, firstly through convenient dissemination processes using the world wide web and secondly by reducing effort required for knowledge reuse by providing concise (or summarized) knowledge extracted from the PPRs which is linked to the specific full PPRs, if required.

This section therefore has demonstrated that TM has useful potential to benefit construction project supply chain design by addressing known problem areas.

7 Conclusion and Future Work

Post Project Reviews (PPRs) are a rich source of knowledge and data for construction supply chains. But dedicated analysis is needed to extract useful knowledge to improve the design the future construction projects. KDT and TM provide mostly automated techniques that aim to discover high level information in huge amounts of PPRs. The focus of this paper has therefore been to determine the potential benefits that TM can offer by automated analysis of PPR to improve the new construction project supply chain design by avoiding mistakes, improving customer service, adopting good practices, making the organization aware of previously known problem areas etc.. A case study of two UK based construction companies has been presented to show the effectiveness of our approach.

The potential for exploitation of this research goes far beyond PPRs since the KDT and text mining should be applicable to a whole range of text based reports and other documents within construction and manufacturing. Future challenges to be addressed relate to the semantic challenges of Synonymy (multiple terms being applied to a single concept) and Term Ambiguity (when a single term is applied to different concepts), where an ontological engineering approach will be adopted in harmony with KDT and TM.

Acknowledgment

We would like to thank our collaborators for the valuable discussions and case study.

References


[2] V. Kumar and N. Viswanadham, “A CBR based decision support system framework for construction supply


