A framework for extracting and representing project knowledge contexts using topic models and dynamic knowledge maps

Xu Jin, Li Zheng, Li Shuliang, Yanyan Zhang

Jin Xu et al., "A framework for extracting and representing project knowledge contexts using topic models and dynamic knowledge maps" Seventh International Conference on Digital Image Processing: ICDIP 2015, edited by Charles M. Falco, Xudong Jiang, Proc. of SPIE Vol. 9631, 96311P

http://dx.doi.org/10.1117/12.2196528

A framework for extracting and representing project knowledge contexts using topic models and dynamic knowledge maps

Jin Xu [1], Zheng Li [1], Shuliang Li [2,1], Yanyan Zhang [1]
School of Economics & Management, Southwest Jiaotong University, Chengdu 610031, China
jinxu@home.swjtu.edu.cn; xj_james@163.com; jeff.lee@my.swjtu.edu.cn
[2] Westminster Business School, University of Westminster, 35 Marylebone Rd, London NW1 5LS, United Kingdom
lish@westminster.ac.uk; shuliangliuk@gmail.com

ABSTRACT

There is still a lack of effective paradigms and tools for analysing and discovering the contents and relationships of project knowledge contexts in the field of project management. In this paper, a new framework for extracting and representing project knowledge contexts using topic models and dynamic knowledge maps under big data environments is proposed and developed. The conceptual paradigm, theoretical underpinning, extended topic model, and illustration examples of the ontology model for project knowledge maps are presented, with further research work envisaged.

Keywords: project knowledge contexts, topic model, dynamic knowledge map, ontology model, project management.

1. Introduction

Project management always involves knowledge. Knowledge itself and associated value are related to a particular context. However, there is a lack of the processing, mining and use of contextual information in the area of project knowledge management. In this "big data" age, with the huge volume of complex information and knowledge in relation to various types of millions of projects across the world and World Wide Web, the needs for appropriate methods and smart tools for locating, mining and interpreting project knowledge contexts have never been greater.

Our study is concerned with project knowledge context extraction and presentation. In order to improve knowledge learning and re-use, our main objectives are to explore how to mine and discover the contents and relationships of project knowledge contexts, and how to symbolize and visualize these.

In this paper, we are going to propose and develop a new framework for extracting and representing project knowledge contexts using topic models and dynamic knowledge maps.

2. Relevant research and key literature review

In this section, we review previous research work on topic models for context discovery, and knowledge maps that integrate contextual information.

2.1 Topic models for context discovery

In their research on wireless applications, Phung et al. [1] utilized probabilistic topic models to extract a higher level of user contexts (e.g., location sequence of user activities) from lower levels. Bao et al. [2] developed a topics-based method for the modelling of mobile service contexts. Pioneering work has also be undertaken by other scholars, typically Zhuang et al. [3] and Zhou et al. [4], to explore topic changes over newspaper and micro-blogs on the Internet, and investigate how to track relevant changes effectively, through the use of time-based, and time-space-based models.

Seventh International Conference on Digital Image Processing: ICDIP 2015, edited by Charles M. Falco, Xudong Jiang, Proc. of SPIE Vol. 9631, 96311P · © 2015 SPIE
CCC code: 0277-786X/15/$18 · doi: 10.1117/12.2196528

Nonetheless, in the literature, there is a lack of attention to the domain characteristic of project documents. Moreover, the potential roles and impacts of project working data and other project information have not been highlighted adequately.

2.2 The knowledge maps incorporating contextual information

According to Eppler [5], a knowledge map contains two parts: a background layer (the contexts and associated inter-relationships), and an individual element layer (experts, project teams, members, etc.). Liu et al. [6] suggested an enterprise contexts-based three layer knowledge map construction approach. Wang et al. [7] proposed a method for representing contextual information for the building industry using ontology models. Chen [11] developed a method and system for representing ontology-based empirical knowledge and performing associated reasoning. Although some researchers have taken into account of circumstances into knowledge maps, there is still a lack of the use of contextual information in the area of project knowledge management.

3. Extension and modification on topic models

In the project management domain, there is great amount of dynamic working data, in addition to the very large volume of project documents. Appropriate methods and tools for dealing with these complex big data are required.
Blei et al. [8] described a method, called Latent Dirichlet Allocation (LDA), a flexible probabilistic model for gathering discrete data. It is assumed that the words of a document come from a mix of topics, each of which has a distribution across the vocabulary. Blei and Lafferty [9] then developed a correlated topic model (CTM) that delivers an effective and natural way of visualizing and investigating unstructured data sets. Wang and McCallum [10] reported an LDA-style topic model, named as Topic Over Time (TOT). It does not reply on Markov assumptions. In their model, each topic is connected to a continuous distribution over timestamps, and for each generated document, the blended distribution over topics is affected by both word co-occurrences and the document's timestamp. The topics' occurrence and correlations change over time.

On the basis of the above-mentioned work [8, 9, 10], we now modify the topic models by making improvements on the following two facets. Firstly, we extend the distributions of words in topics of the topic models for working data, which emphasizes the influence and impacts of contextual information on word distributions over topics. Secondly, the relationships between project knowledge contexts can be determined through the setting and formulation of topic distributions. Different topic distributions can generate different project knowledge contexts with different relationships. Our extended topic model is shown in Fig.1, where empty circles denoting implicit or latent variables; shaded solid circles standing for observable variables or observations; P –phase or stage of a project; M – the number of documents (corpus); N - the number of words in a document; W – a word in a document; z – the topic that a word belongs to; θ – the probability of a topic in a document; β – the probability distribution of each topic's corresponding words in the word list or vocabulary. It is worthy of note that the words in project working data can affect β. It is also important to mention that phased project working data can influence prior probability distribution of the topics of project documents.

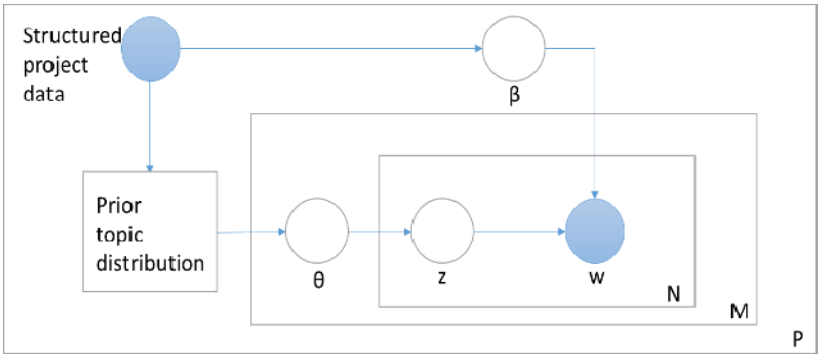

Fig.1. The extended topic model

The processing procedure of topic extraction for project knowledge contexts, using time-related topic distribution is illustrated in Fig.2. It should be mentioned that the inclusion and consideration of relevant information about working data will enable the obtained topic relationships to be more accurate and meaningful. If project documents are used for the purpose of extraction, then the original topics are obtained. If project working data are added and considered for extraction, then the modified topics are acquired.

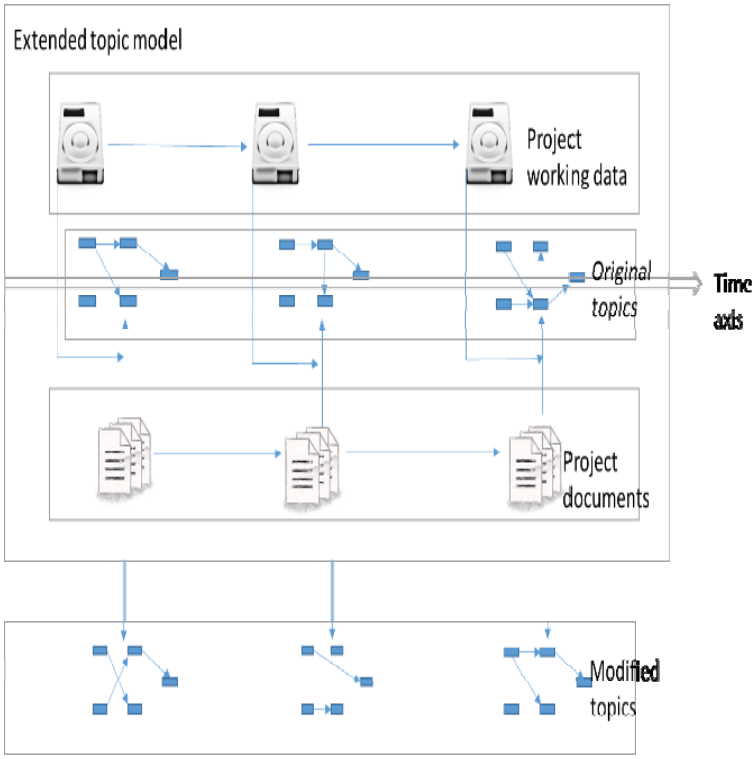

Fig.2. The processing procedure of topic extraction for project knowledge contexts, using time-related topic distributions

4. Project knowledge map construction method considering project knowledge contexts

It is essential to clarify project knowledge activities, project knowledge, contexts, and the relationships amongst contexts. In particular, the relationships can be beforehand, afterward, inclusion, and/or causal, etc. On the basis of these, we can connect and link knowledge, contexts, and relationships amongst contexts.

In order to create and construct a knowledge map, the ontology technique can be employed to specify a shared conceptual model, define commonly accepted vocabularies and their relationships in a domain, and then formalize relevant concepts and relationship networks.

An ontology model consists of classes, properties, and individuals. The ontology modeling tool, Protégé, designed by Stanford University, is employed in our study. We apply OWL (Web Ontology Language) to the definition and representation of project knowledge, contexts and associated relationships.

An example of defining a class in an ontology model of a project knowledge map is given below.

```
<owl:Class rdf:about="&Project-ontology;ProDocument">
```

```
        <rdfs:comment>Documents related to the project, such as project tender documents, table of
cost budget, table of surplus analysis, the quality schedule and so on</rdfs:comment>
        <rdfs:subClassOf rdf:resource="&Project-ontology;DomainConcept"/>
        <rdfs:subClassOf>
          <owl:Restriction>
            <owl:onProperty rdf:resource="&Project-ontology;hasDecrition"/>
            <owl:someValuesFrom rdf:resource="&xsd;string"/>
          </owl:Restriction>
        </rdfs:subClassOf>
        <rdfs:subClassOf>
          <owl:Restriction>
            <owl:onProperty rdf:resource="&Project-ontology;isCreatedInProject"/>
            <owl:someValuesFrom rdf:resource="&Project-ontology;Project"/>
          </owl:Restriction>
        </rdfs:subClassOf>
          ……
    </owl:Class>
```

The specification of a property for an ontology model, using milestone as an example, is provided as follows.

```
    <owl:ObjectProperty rdf:about="&project-ontology;hasMilepostActivity">
      <rdfs:range rdf:resource="&project-ontology;MilepostActivity"/>
      <rdfs:domain rdf:resource="&project-ontology;ProWBS"/>
      <rdfs:subPropertyOf rdf:resource="&project-ontology;hasWBSActivityValue"/>
          ......
     </owl:ObjectProperty>
```

The definition of an individual for an ontology model of project knowledge map is demonstrated below.

```
    <owl:NamedIndividual rdf:resource="&Project-ontology;ProWBSActivity"/>
      <hasItem rdf:datatype="&xsd;double">2.1</hasItem>
      <isBelongToProWBS rdf:resource="&Project-ontology;WBS_of_ERP_project_of_X_enterprise"/>
      <hasChecked rdf:resource="&Project-ontology;Yes"/>
    ......
    </owl:NamedIndividual>
```

5. Research hypotheses

On the basis of the work discussed in the previous sections, the authors' own reflection, the following hypotheses are formulated for further research.

Hypothesis 1. Our proposed framework integrating topic models and dynamic knowledge maps for extracting and representing project knowledge contexts will help improve the sharing of project knowledge.

Hypothesis 2. Our proposed framework will help improve the re-use of project knowledge.

Hypothesis 3. Our proposed framework will help the users' learning on project knowledge.

Hypothesis 4. Our proposed framework will help improve the quality of projects.

Hypothesis 5. Our proposed framework will help reduce the costs for projects.

Hypothesis 6. Our proposed framework will help improve project completion time.

6. Conclusions and further research

This paper has been sought to propose and develop a new framework for the extraction and representation of project knowledge contexts from a big data perspective.

The scope of the topic models have been broadened. The distributions of words in topics of the topic models has been extended to deal with working data, and highlight the influence and impacts of contextual information on word distribution over topics. The relationships between project knowledge contexts can now be determined by formulating topic distributions.

In addition, dynamic knowledge maps have been incorporated into the paradigm. Relevant examples of the specification of classes, properties and individuals in an ontology model for project knowledge maps considering contextual information are provided and illustrated. Research hypotheses have also been formulated.

Further work is being undertaken to realize, implement and validate the proposed framework, and test the associated research hypotheses.

Acknowledgments

This study is supported by grants from National Natural Science Foundation of China (Grant numbers: 71472158 & 71002064), the Fundamental Research Fund for the Central Universities of China (2682013CX075), the Soft Science Research Programme of Sichuan Province (2013ZR0043), the University of Westminster staff research allowances (UK), and Sichuan 100-Talent Scheme grant (Sichuan bai ren ji hua, China. Grant number: Chuan Zu Tong [2012] 142).

References

[1] Phung, D., Adams, B., Venkatesh, S. and Kumar, M., "Unsupervised context detection using wireless signals", Pervasive and Mobile Computing, 5(6), 714–733 (2009).

[2] Bao, T., Cao, H., Chen, E., Tian, J. and Xiong, H., "An unsupervised approach to modeling personalized contexts of mobile users", Knowledge and Information Systems, 31(2), 345–370 (2012).

[3] Zhuang, T., Wang, P. and Cheng, Q., "A time-context dependent method for micro-blogging topic extraction" (In Chinese), Information Resources Management Journal, (3), 40-46 (2013).

[4] Zhou, Y. and Du, J., "Time-space context model-based topic tracking", Journal of Huanan University of Science & Technology (Natural Science), (8), 82-87 (2012).
[5] Eppler, M. J., "Making knowledge visible through intranet knowledge maps: concepts, elements, cases", Proceedings of the 34th Annual Hawaii International Conference on System Sciences, 9–18 (2001).
[6] Liu, L., Li, J. and Lv, C., "A method for enterprise knowledge map construction based on social classification", Systems Research & Behavioral Science, 26(2), 143–153 (2009).
[7] Wang, H., Boukamp, F. and Elghamrawy, T., "Ontology-based approach to context representation and reasoning for managing context-sensitive construction information", Journal of Computing in Civil Engineering, 25(5), 331–346 (2011).
[8] Blei, D., Ng, A. and Jordan, M., "Latent dirichlet allocation", Journal of Machine Learning research, 3, 993-1022 (2003).
[9] Blei, D. and Lafferty, J., "Correlated topic models", Advances in Neural Information Processing Systems, 18, 147 (2006).
[10] Wang, X. and McCallum, A., "Topics over time: a non-Markov continuous-time model of topical trends", Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, New York, USA, 424-433 (2006).
[11] Chen, Y., "Development of a method for ontology-based empirical knowledge representation and reasoning", Decision Support Systems, 50(1), 1-20 (2010).