A Review Report on Requirements Analysis with Data Mining

Faizan Berlas 1

1Virtual University of Pakistan

December 7, 2023

Abstract

Software requirements engineering is the process of eliciting, analyzing, negotiating, specifying, documenting, and validating software requirements. These requirements are gathered from user needs, stakeholder needs, domain information, existing system information, organizational standards, and external regulations. Traditionally, the requirement engineering process has been driven by stakeholders’ needs. However, the advent of the Internet of Things (IoT) devices, mobile devices, social networks, and other big data technologies have caused a shift from the traditional mode of the requirements acquisition process to new data-driven and automatic modes of requirement elicitation by utilizing data mining techniques. The traditional methodologies of requirement engineering processes are inadequate for modern-day software applications due to their inability to effectively capture user requirements from the ever-increasing volume of data, customer requirements, and feedback. Consequently, the utilization of data mining techniques is gaining popularity in the field of requirement engineering. These techniques are employed to extract valuable insights from data and elicit customer requirements, providing a more robust approach to addressing the challenges posed by modern software development. This research paper comprehensively reviews various data mining techniques and methodologies used for gathering, eliciting, and analyzing software requirements. It offers valuable insights into the significant areas of data mining, requirements analysis, and elicitation that have undergone extensive research in recent years. By providing a thorough understanding of requirements engineering with data mining, the paper serves as a valuable resource for researchers venturing into the realm of requirement analysis and elicitation with data mining and seeking potential avenues for future research.
A Review Report on Requirements Analysis with Data Mining

Muhammad Faizan Berlas  
Department of Computer Science  
Virtual University of Pakistan  
Lahore, Pakistan  
ms220400075mfb@vu.edu.pk

Abstract—Software requirements engineering is the process of eliciting, analyzing, negotiating, specifying, documenting, and validating software requirements. These requirements are gathered from user needs, stakeholder needs, domain information, existing system information, organizational standards, and external regulations. Traditionally, the requirement engineering process has been driven by stakeholders’ needs. However, the advent of the Internet of Things (IoT) devices, mobile devices, social networks, and other big data technologies have caused a shift from the traditional mode of the requirements acquisition process to a new data-driven and automatic mode of requirement elicitation by utilizing data mining techniques. The traditional methodologies of requirement engineering processes are inadequate for modern-day software applications due to their inability to effectively capture user requirements from the ever-increasing volume of data, customer requirements, and feedback. Consequently, the utilization of data mining techniques is gaining popularity in the field of requirement engineering. These techniques are employed to extract valuable insights from data and elicit customer requirements, providing a more robust approach to addressing the challenges posed by modern software development. This research paper comprehensively reviews various data mining techniques and methodologies used for gathering, eliciting, and analyzing software requirements. It offers valuable insights into the significant areas of data mining, requirements analysis, and elicitation that have undergone extensive research in recent years. By providing a thorough understanding of requirements engineering with data mining, the paper serves as a valuable resource for researchers venturing into the realm of requirement analysis and elicitation with data mining and seeking potential avenues for future research.

Keywords—Data mining, Requirement engineering, Requirements elicitation, Requirement analysis.

I. INTRODUCTION

Requirement analysis, also known as requirements engineering or requirements gathering, is a systematic process for identifying, documenting, and validating the needs and expectations of stakeholders for a software system or product. It includes understanding the system's goals, functions, and constraints, as well as eliciting, analyzing, and prioritizing requirements to ensure that the final solution meets the desired objectives [6]. The advancements in cloud computing, industrial internet, machine learning, and other emerging technologies have caused a huge influx of data from a variety of heterogeneous sources [4]. Because of the huge amount of heterogeneous data, classic methodologies of the software requirements analysis processes fail to capture efficiently capture customer requirements for modern-day software systems. As a result, traditional modes of requirements engineering processes are being replaced by data-driven, knowledge-based, and autonomous requirement engineering processes that are backed by data mining techniques and methodologies [1].

In this paper, several publications are reviewed to provide an overview of the data mining techniques that are used for requirement analysis, and elicitation. The surveyed publications include published literature from 2008 to 2023.

A variety of data mining techniques have been researched for software requirement engineering. [4], [11] and [12] utilized Latent Dirichlet allocation (LDA) topic modeling to capture software requirements from big data systems. [1], [2], and [6] have employed natural language processing, sentiment analysis, and other data mining techniques for gathering software requirements from online data sources.

Data mining is used for analyzing online reviews and extracting requirements for continuous product improvement. Zhao et al. [9] introduce an innovative approach for dynamically extracting customer requirements by analyzing the evolving customer reviews related to products and their attributes. To assess the effectiveness and feasibility of their proposed method, [9] conducted a case study on multi-generation phone products, utilizing online reviews as the primary data source.

Data mining techniques have also been employed for gathering the non-functional requirements of modern-day software systems. Portugal [10] proposes a knowledge-based semi-automated process for mining keywords to gather non-functional requirements.

This paper is assembled as follows: Section II presents the research work that was collected and analyzed for the literature review. Section III provides the findings from numerous research papers and discusses them in the context of data mining for requirements analysis and elicitation. Finally, section IV presents the conclusions, the limitations of the current research work, and ideas for future research.

II. RELATED WORK

To conduct a comprehensive literature review, a diverse collection of research publications covering various data mining methodologies, and techniques for requirement engineering and elicitation was gathered. These publications were sourced from reputable online research databases, with a specific focus on selecting journals indexed in Scopus, Science Citation Index Expanded (SCIE), or Emerging Sources Citation Index (ESCI) for their rigorous peer-review process and academic authority.

Table I shows the list of 12 papers that were collected and presents them in ascending order by the year of publication. For each paper, the area of research work is presented along with a brief description of the research area and the proposed approach provided in the paper.
TABLE I. RELATED WORK IN DATA MINING FOR REQUIREMENTS ENGINEERING

<table>
<thead>
<tr>
<th>Paper</th>
<th>Year</th>
<th>Research Area</th>
<th>Proposed Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Britos et al. [8]</td>
<td>2008</td>
<td>Data mining for business intelligence project requirements</td>
<td>Case study</td>
</tr>
<tr>
<td>Castro-Herrera et al. [2]</td>
<td>2008</td>
<td>Data mining and recommender technology for requirement elicitation</td>
<td>Recommender system framework utilizing unsupervised learning techniques</td>
</tr>
<tr>
<td>Dong et al. [6]</td>
<td>2010</td>
<td>Text data mining for requirements gathering</td>
<td>Conjoint analysis based on vector model, ideal-point model, and part-worth model</td>
</tr>
<tr>
<td>Zhang et al. [13]</td>
<td>2014</td>
<td>Mining online reviews for gathering requirements</td>
<td>Systematic mapping study</td>
</tr>
<tr>
<td>Portugal et al. [10]</td>
<td>2018</td>
<td>Data mining for non-functional requirements elicitation</td>
<td>Natural language processing (NLP) techniques</td>
</tr>
<tr>
<td>Alwadain and Alshargi [1]</td>
<td>2019</td>
<td>Data mining for requirements elicitation from user feedback and comments</td>
<td>Classification, topic modeling, and sentiment extraction</td>
</tr>
<tr>
<td>Lim et al. [11]</td>
<td>2021</td>
<td>Data mining for requirements elicitation</td>
<td>Sentiment analysis</td>
</tr>
<tr>
<td>Zhao et al. [9]</td>
<td>2022</td>
<td>Data mining for requirements elicitation</td>
<td>Deep transfer learning and improved Latent Dirichlet allocation (ILDA)</td>
</tr>
<tr>
<td>Wu et al. [12]</td>
<td>2022</td>
<td>Data mining for requirements elicitation</td>
<td>Empirical study</td>
</tr>
<tr>
<td>Hao et al. [4]</td>
<td>2023</td>
<td>Data mining for requirements elicitation</td>
<td></td>
</tr>
</tbody>
</table>

III. DISCUSSION

This section provides the findings and proposed approaches presented in the research work that was selected for the literature review. The research questions are designed to establish a relationship between the published research work and data mining techniques for requirement analysis, engineering, and elicitation.

A. RQ1: How can data mining be used for requirements gathering, analysis, engineering, and elicitation?

With the help of data mining, valuable information can be extracted from big data. This information can be analyzed for gathering customer requirements. The information can be in the form of surveys, customer reviews, feedback, or comments. This research question offers a comprehensive view of the chosen research papers that explore the utilization of data mining in requirements analysis and elicitation.

Zhao et al. [9] propose a new method for dynamically mining customer requirements by analyzing the evolving customer reviews about products and their attributes in online reviews. This helped in understanding the customer needs and provide valuable insights for future product improvement. The method proposed by [9] consists of the following three steps;

1) Employing text mining techniques to collect online review data for multiple generations of products and identify specific product attributes.
2) Calculating attention and sentiment scores for the product attributes using natural language processing tools and integrating them with corresponding satisfaction scores.
3) Determining the improvement direction for future product iterations based on the changing satisfaction scores of product attributes.

To validate the effectiveness and practicality of the proposed approach, [9] also conducted a case study on multi-generation phone products using online reviews.

Lim et al. [11] provide a systematic literature review of the modern methods used in automatic requirement elicitation for developing information systems. Automatic requirement elicitation is a process of extracting requirements automatically from the given data. An information system is an integrated set of components for collecting, storing, and processing data and provides information, knowledge, and digital products. The literature review by [11] uses the published research work that focuses on three types of dynamic data sources for eliciting requirements for information systems. Dynamic data sources include data that is continuously changing. The three types of dynamic data sources are human-sourced, process-mediated, and machine-generated. The process of eliciting requirements based on various data sources is called data-driven requirements elicitation. [11] attempted to find answers to the following research questions;

1) Types of dynamic data used for automated requirement elicitation?
2) Techniques and technologies used for automating requirement elicitation?
3) Outcomes of requirement elicitation automation?

To find answers to the research questions, [11] devised research article inclusion criteria and various search strategies and collect 68 research papers for analysis. The selection criteria used by [11] consist of the following factors;

1) Automated requirement elicitation
2) Requirement elicitation from digital and dynamic data sources
3) Digital and dynamic data sources created without the intention of eliciting requirements
4) Change in requirements must include elicitation of new requirements.
5) Research articles are peer-reviewed
6) Research articles are written in English

[11] searched comprehensively in six electronic databases to find relevant articles for the systematic literature review. These databases include Scopus, Web of Science, ACM Digital Library, IEEE Xplore, EBSCOhost, and ProQuest.

After finalizing the research articles and respective research questions, [11] developed and refined an analytical framework for answering the research questions. For this purpose, the following information was extricated from the systematic literature review:

1) Types of dynamic data sources used for automating requirements elicitation
2) Specific types of dynamic data used for automating requirements elicitation
3) Integration of data sources
4) Relation of dynamic data to a given organization
5) Additional domain knowledge
6) Techniques used for requirement elicitation automation, including process patterns for automating requirements elicitation;
7) Use of aggregation in automating requirements elicitation
8) Use of visualization in automating requirements elicitation
9) Methods for evaluation requirement elicitation automation
10) Types of requirements generated by automating requirement elicitation.

11) Additional requirement engineering process activity supported by the automation process

The work of [11] is of significant importance because it provides an overview of state-of-the-art data-driven requirements elicitation techniques and is the first systematic literature review that focuses on automating requirement elicitation from big data sources including both human-sourced and machine-generated data.

B. RQ2: What are the data mining techniques that are used for requirements gathering and elicitation?

This question attempts to determine the types of data mining techniques that have been used in published literature for gathering, analyzing, and eliciting software requirements in modern-day software systems. In this context, the most widely used data mining techniques for analyzing requirements are topic modeling, natural language processing, and sentiment analysis.

IV. CONCLUSION AND FUTURE WORK

This research paper provides an overview of data mining techniques that are used for requirements analysis and elicitation. A literature review of published research papers has been carried out to find how data mining is used for requirement analysis and what types of data mining techniques are used for software requirements engineering. As per the findings of this research, the data mining techniques and models that are used for requirements analysis include sentiment analysis, sentiment extraction, Latent Dirichlet allocation (LDA) model, deep learning, topic modeling, natural language processing, Kano model, knowledge-based search, vector space model, and unsupervised learning.

There is a lot of published literature on requirement analysis and elicitation from data mining; however, most of the research work is based on theoretical assumptions. The validity of the findings published in the research work is yet to be established and there is a need for detailed empirical studies for evaluating the usefulness of the proposed techniques in real-life business and industrial problems. Published literature on the utilization of data mining for creating health and education software is rare. Similarly, there are very few published case studies describing the benefits and competitive advantages of data mining for developing highly desired software for business and marketing.

One of the limitations of this research paper is the small number of published research papers for the literature review. This is because only those research papers that have open access and exist in Scopus, Science Citation Index Expanded (SCIE), or Emerging Sources Citation Index (ESCI) indexed journals were selected. Moreover, this literature review tries to combine various areas of data mining and software requirements elicitation in a single article i.e. empirical studies, machine learning techniques, and systematic literature reviews for data mining and requirements elicitation, are all combined in this research work. The utilization of data mining techniques in gathering, analyzing, hence eliciting requirements is a rapidly expanding field with vast implications and there needs to be a more systematic approach for conducting its literature review. Reviewing the published literature based on specific elimination and inclusion criteria and focusing on any particular aspect of data mining and requirements elicitation will provide a more focused overview of the subject.

REFERENCES


