Process Mining in Knowledge Maintenance A Case Study

Ming Li

School of Business Administration, China University of Petroleum, Beijing 102249, China, limingzyq@gmail.com

Abstract

The quality of knowledge in the knowledge repository determines the effects of knowledge reusing and sharing. In knowledge management systems (KMS), knowledge maintenance events are recorded in process logs. In order to take knowledge maintenance process logs to discover process, control, organizational, and social structure and construct a more appropriate knowledge maintenance process model, process mining is applied to the knowledge maintenance. The paper demonstrates the applicability of process mining using a real case of a knowledge maintenance process in an aviation design institute. In this paper, we analyzed the knowledge maintenance from two different perspectives: (1) the process perspective, which is used to find a good characterization of knowledge maintenance tasks and paths and (2) the organizational perspective which is used to find relations between individual performers. The results show that process mining can be used to provide new insights that facilitate the improvement of existing knowledge maintenance process.

Keywords: Knowledge Maintenance, Knowledge Management, Process Mining, Workflow Mining

1. Introduction

With the development of knowledge-based economy, comprehension of knowledge management is deepening gradually [1]. Many enterprises have launched enterprise knowledge management plans [2, 3]. To be the foundation of Knowledge Management, knowledge management system (KMS) is the key to implement the knowledge management strategy [4]. Enterprises are attempting to leverage their knowledge resources by employing knowledge management systems [5], a key component of which is the knowledge repository. The quality of the knowledge in the knowledge repository determines the effects of knowledge reusing and sharing [6]. If the knowledge in the knowledge repository contains irrelevant information, the knowledge retrieval and knowledge sharing will be interfered. Even worse, some of the knowledge is incorrect and will inevitably mislead future decision makings and operations for an organization. Without a control of knowledge maintenance process, the quality of knowledge cannot be guaranteed, so it is necessary to control knowledge maintenance process, such as the knowledge acquisition process, the knowledge removal process and so on.

Workflow technique has been introduced into the knowledge acquisition process, which is used for ensuring the quality of knowledge acquisition [7]. The knowledge maintenance processes are not under complete control of the system to improve the flexibility. Predefined tasks cannot be skipped but the additional task can be added manually by the process originator to meet the unofficial needs of knowledge maintenance process change. Moreover, the authorized people can delegate the tasks to others unofficially who are considered to be qualified to perform these tasks. All these lead to the discrepancy between the predefined knowledge maintenance process model and the actual knowledge maintenance process model.

Knowledge management systems with embedded workflow system provide knowledge maintenance process logs. Typically such logs register the process instances. Every instance comprises the tasks in the process and the timestamp, performer and some additional data of each completed task. The knowledge management systems allow for a lot of flexibility. The logs reflect the actual knowledge maintenance process but not mere the predefined process model.

Process mining techniques can be used to support the redesign of the knowledge maintenance process model by analyzing the logs [8]. It has already been successfully applied in many areas such as the provincial office [9], multi-agent system [10], et al. These applications all focus on the business process. Nevertheless, only a few studies have been conducted on the application to the
Process Mining in Knowledge Maintenance: A Case Study
Ming Li

knowledge maintenance. For example, we have proposed the knowledge maintenance process mining algorithms [8]. In order to know how process mining can be used for obtaining insights related to knowledge maintenance process, in this paper, we make the further study on the knowledge maintenance process mining and focus on the applicability of process mining in knowledge maintenance.

In the paper, we describe a case study based on the application of process mining to knowledge maintenance in an aviation design institute of China. The knowledge management system has already been implemented in the institute. It is doubt whether the current process model is appropriate. The institute is urgent to obtain the actual knowledge maintenance process model and search for opportunities to optimize the model. We will elaborate how process mining is used for obtaining insights related to knowledge maintenance process from two perspectives (1) process perspective and (2) organization perspective.

The rest of the paper is organized as follows: Section 2 provides an overview of process mining. Section 3 shows the case study in more detail, along with our analysis and findings. Finally, Section 4 concludes the paper.

2. Process mining

Process mining is applicable to a wide range of systems. The only requirement is that the system produces event logs, thus recording the actual behavior. Most knowledge management systems, especially those with the embedded workflow system, provide knowledge maintenance process logs. Typically such logs register the process instances. Every instance comprises the tasks in the process and the timestamp, performer and some additional data of each completed task.

The goal of process mining is to extract information from these logs. In practice, there is often a significant gap between what is prescribed or supposed to happen, and what actually happens. Only a concise assessment of reality, which process mining strives to deliver, can help in verifying process models, and ultimately be used in system or process redesign efforts [11]. In organizations, the actual knowledge maintenance processes and the processes perceived by the management are not fully identical.

The idea of process mining is to discover, monitor and improve real processes by extracting knowledge from event logs. We consider two basic types of process mining: (1) discovery and (2) conformance.

**Discovery.** Traditionally, process mining has been focusing on deriving information about the original process model and the organizational context from enactment logs. Process mining is not restricted to the process perspective and also includes other perspectives such as the organizational. The process perspective is concerned with the control flow, i.e., the causal ordering of activities. The goal is to derive a process model. The organizational perspective is concerned with the organizational structure and the people within the organizational units involved in the process. [8] focuses on the knowledge maintenance process mining and proposed the corresponding process mining and organization mining algorithm.

**Conformance.** There is an a-priori model. This model is used to check if reality conforms to the model. Conformance checking may be used to detect deviations, to locate and explain these deviations, and to measure the severity of these deviations [12].

It is important to note that process mining should not require all possible observations to be present in the log. In general only fraction of the possible behavior will actually be observed.

3. Knowledge Maintenance Process Mining

In the study we analyzed the process logs generated by the knowledge management system (KMS) that had been implemented in an aviation design institute which designed warplanes and civil aircrafts. Note that the information involving the secret are not covered in the case study according to the secrecy requirement. Nevertheless it will not affect the illustration and the understanding of the following case study.

Firstly, we give a brief overview of the institute and the implemented knowledge management system. The formal organizational model of the institute is made up of seventeen departments and
thirty-four units. Figure 1 shows part of the organizational model. The units are branches of the department. All the designers except the leaders in the department are located in specific units. There are nine vice chief designers and each designer is responsible for one or two departments in the institute.

![Organizational Model](image)

**Figure 1.** Formal organization model of the institute

The state has begun to develop the aviation industry greatly and the number of aviation design projects has been increased a lot in the institute. The designers complained that the time spent on knowledge retrieval was too much and there was a lack of methods to retrieve knowledge more efficiently. Moreover, more and more crucial knowledge was lost because of the experts’ retirement, and there was no approach to collect the experience knowledge systemically. Therefore, they were in urgent need of knowledge management system.

The institute has strict requirements for the quality of knowledge in the repository. Even few errors in the knowledge may cause a disaster. It is necessary but difficult to construct an appropriate knowledge maintenance process model. If the process is simple, the quality may not be guaranteed. If the process is complex, the knowledge in the repository cannot be updated timely and the effect of knowledge sharing may be reduced.

Knowledge maintenance processes were jointly designed by leaders of the institute and the KMS designers. The knowledge maintenance includes two kinds of processes. One is the acquisition process and the other is the archive process. The acquisition process ensures the quality of knowledge entering the repository for knowledge sharing. The archive process ensures that the knowledge to be archived will not be used frequently in the future. The knowledge in the repository cannot be removed because it would be used when the former designed aviations according to it needs repair. Therefore, the outdated knowledge is archived in the repository and can not be found in normal knowledge retrieval afterwards.

In the beginning they thought that the more tasks in the process the more errors could be found in the knowledge. The knowledge maintenance processes contained all the tasks. With the implementation of the KMS they found that the knowledge maintenance process model could not meet the actual knowledge maintenance needs. There was a lot of knowledge which had been confirmed by the state, such as standard knowledge, and only needed to be confirmed by the vice chief designer. But the unit leader and the department leader had to check it according to the predefined knowledge maintenance process model. The knowledge maintenance efficiency decreased and the leaders took on heavy burden owing to the redundant tasks.

Afterwards the knowledge management system was upgraded. Each repository has its knowledge maintenance processes. For example, the acquisition process of airworthiness knowledge consists of two tasks. However, the acquisition process of lessons learned knowledge consists of four tasks. In addition, the predefined tasks cannot be skipped but the additional task can be added manually by the process originator in order to further improve flexibility.

Most knowledge maintenance tasks have to be completed by leaders and the vice chief designers because of their profound knowledge, but they are too busy in designing the aviation to execute the tasks. In order to resolve the conflict, the task delegation function is provided in the KMS. The authorized designers can delegate the tasks to the designers unofficially who are considered to be qualified to perform the tasks. However, the actual knowledge maintenance organization model cannot be obtained because the delegations are revoked irregularly.

It just so happens that the institute has been enforcing the secrecy rules. The knowledge maintenance process has not been reconfigured officially. In order to keep on maintaining the knowledge some departments made the unofficial rules that designers in the department should add the secret examination task in the knowledge maintenance process manually. The knowledge
maintenance processes would be rejected if the knowledge was not examined from the secrecy perspective.

There are discrepancies between the actual process and the process as perceived by the institute. The institute needed to get a deeper understanding of the process and found opportunities for optimization. We gathered the knowledge maintenance process logs in the knowledge management system in eight months and applied the proposed process mining methods to analyzing these logs. The analysis aims to get the actual knowledge maintenance model from the three perspectives: (1) process perspective and (2) organization perspective. The results of the analysis will be the reference to the optimization of the knowledge maintenance model. Given the scope of this paper, we focus on the representative but not all the mining results.

3.1 Preprocessing of Logs

Knowledge maintenance tasks are performed according to the predefined process model. The generated process logs to be used for process mining are stored in the log repository. Since the logged data in the log repository cannot be directly used in process mining, we have to preprocess process logs. The process instances that have been accomplished in the log are distilled and used for process mining. Moreover, the raw data should be converted into a suitable format. In dependency mining some part of the raw data should be elicited for the raw data includes more fields than required. But in the organization mining, the data that is relevant to the knowledge itself would be incorporated into the raw data. We give a fraction of the log as shown in Table 1. The process logs were generated by the knowledge management system (KMS) that had been implemented in an aviation design institute. Each of the process instances includes the task name, the corresponding performers and sequence of the tasks. When mining from the process perspective, we may eliminate the performers of the log as we only concentrate on the tasks and the relation between them. Also, when mining from the organization perspective, we only reserve the performers and the relations in the log for we focus on the relation of the performers.

<table>
<thead>
<tr>
<th>Id</th>
<th>Logs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Start, Jack), (Exam, Bill), (Start, Jack), (Exam, Bill), (Secrecy, Eileen), (Censor, Mona), (Start, Jack), (Exam, Bill), (Secrecy, Eileen), (Censor, Mona), (Approve, Robert), (Check, Pete)</td>
</tr>
<tr>
<td>2</td>
<td>(Start, Shirley), (Exam, Bill), (Start, Shirley), (Exam, Bill), (Secrecy, Eileen), (Censor, Mona), (Start, Shirley), (Exam, Bill), (Secrecy, Eileen), (Censor, Mona), (Approve, Robert), (Check, Pete)</td>
</tr>
<tr>
<td>3</td>
<td>(Start, Shirley), (Exam, Bill), (Start, Shirley), (Exam, Bill), (Secrecy, George), (Censor, Mona), (Start, Shirley), (Exam, Bill), (Secrecy, George), (Censor, Mona), (Approve, Fred), (Check, Pete)</td>
</tr>
<tr>
<td>4</td>
<td>(Start, Jack), (Exam, Bill), (Secrecy, George), (Start, Jack), (Exam, Bill), (Secrecy, George), (Censor, Mona), (Approve, Fred), (Check, Pete)</td>
</tr>
<tr>
<td>5</td>
<td>(Start, BOB), (Exam, Bill), (Secrecy, George), (Start, Bob), (Exam, Bill), (Secrecy, George), (Censor, Mona), (Approve, Fred), (Check, Pete)</td>
</tr>
<tr>
<td>6</td>
<td>(Start, Jack), (Exam, Bill), (Secrecy, George), (Censor, Mona), (Approve, Fred), (Start, Jack), (Exam, Bill), (Secrecy, George), (Censor, Mona), (Approve, Fred), (Check, Pete)</td>
</tr>
</tbody>
</table>

3.2 Mining the process perspective

In this section, we focus on the task and dependency mining. We take the lessons learned knowledge acquisition process as an example to illustrate the mining from the process perspective [8]. At first we introduce the original process model as shown in Figure 2.

![Figure 2. Original process model of the lessons learned knowledge acquisition](image-url)
In figure 2 there are five real vertexes which are represented by the circles with solid edge and one virtual vertex which is represented by the circle with dotted edge. The virtual vertex only means that all the tasks in the process are completed and the knowledge has entered the repository.

The meanings of the real vertexes are clarified as follows.

Vertex Start is the start task of the process. In other words, the process is originated.

Vertex Exam means the examination task which is the basic quality examination.

Vertex Censor means the censoring task which is the quality examination from the specialty aspect.

Vertex Approve means the approving task which is the quality examination from the aviation model aspect. It is often performed by the vice chief designer.

Vertex Check is the last task which is performed by the chief designer. The documents that are relative to the aviation should be checked by the chief designer traditionally.

We used the knowledge maintenance process mining algorithm to analyze the lessons learned knowledge acquisition process log. Figure 3 shows the mining result.

Compared with the original process model, the vertex Secrecy is added and the vertex Check is removed. The added vertex Secrecy means the secrecy control task which is the secrecy related examination according to the secrecy rules. The contents of some tasks are not identical to the former tasks. The censoring task has an additional content which is to determine the scope of the knowledge sharing.

The performer of the check task is the chief designer. He is responsible for all the designs of the aviation and the related information systems in the institute. Little time is left for him to check the knowledge. As a whole, he gave the results according to suggestions given by the vice chief designer. The check task takes no effects and is removed from the mined process model.

Although the mining result shows that the secrecy control task and censoring task are performed in parallel, in each specific department they are performed in series. The sequence of the two tasks varies with the department for the lack of official rules. We were not limited to one specific department but focused on the whole institute in our analysis. Therefore, the sequence of the two tasks in the mining result is in parallel.

From the above illustration we see that there were discrepancies indeed. The constructed knowledge maintenance process model was inevitably influenced by the other business processes. It is very important to get the actual process model and optimize the predefined process model to improve the knowledge maintenance process efficiency.

In the remainder, we no longer focus on the process perspective and direct our attention towards the organization perspective.

3.3 Mining the organization perspective

In this section, we examine the mining from the organization perspective. Specifically the organization mining includes two parts that are social network analysis and knowledge maintenance organization model mining. In the section we used the distilled log related to the General Department and the Equipment Department to construct the social network and generate an organizational model respectively.

To analyze the relationship between performers, we performed social network analysis. We used the handover of knowledge metric to derive it. Figure 4 shows the social network generated from the distilled log.
There are forty-one nodes in the social network and each node represents a user. A directed arc between B66 and B61 represents that on some occasion B66 handed knowledge over to B61.

We can find core users in the social network using the centralization degree. Besides Z1 there are another three people with higher centralization degree who are K1, K2 and B32. These people are considered as the core people in the knowledge maintenance process. More users handed knowledge over to them.

![Figure 4. Social network in the General Department and the Equipment Department](image)

In the figure 4, most users are connected and there is only one user B68 isolated from others. It can be easily deduced that most users joined in the knowledge maintenance process and only B68 was inactive. B68 handed no knowledge over to others and no knowledge was handed over to him either. B68 was a novice in the institute. He was afraid of making mistakes so he preferred not to originate the knowledge maintenance process. In addition, he did not have enough knowledge to perform the knowledge maintenance tasks indeed.

The nodes without income degree such as B27 mean that the users only originated the knowledge maintenance process. They were not qualified to perform other knowledge maintenance tasks. Usually they were ordinary designers.

There are two clusters in the social network which are formed by the designers of the general department at the upper part of the figure and the Equipment Department in the lower part of the figure respectively. It reflects that the knowledge maintenance tasks were performed within the department. There is no relation between the two departments. The knowledge in the Ground Equipment Unit can only be maintained by the Ground Equipment Unit and its senior department that is the Equipment Department. The General Department is not qualified to maintain the equipment knowledge. It is reasonable that each department in the institute is in a specific knowledge area.

The connecter of the two clusters is Z1. He was involved in knowledge maintenance processes of the two departments as shown in Figure 4. Since he is the vice chief designer, he is responsible for the two departments. In the same way the department leader may be involved in knowledge maintenance processes of the subordinate units.

The density of the two clusters is different. The density of the lower cluster which represents the Equipment Department is higher than the upper cluster which is the General Department.

The density is determined by connections between users that are derived from the handover of the knowledge. The more users are involved in the handover process, the higher the density is. The following are the main factors that influence the density.

The first two factors are the number of kinds of knowledge maintenance process and the number of tasks in the process, which are positively associated with the amount of handover of knowledge between the users. The number of the tasks that each user is involved in and the number of performers in each task also have close correlations with the density. There will be more relations between the users in the successive tasks with the knowledge handed over to more users.

In the distilled log, we found that there was no difference of knowledge maintenance processes between the two departments. We inferred that the difference of density was mainly caused by the last two factors. That is, more users were involved in some tasks in the Equipment Department. The following actual knowledge maintenance organization model will verify the inference.
The knowledge maintenance organization model mining algorithm is used to analyze the distilled logs. Figure 6 shows the mined actual knowledge maintenance organization model and figure 5 shows the predefined knowledge maintenance organization model respectively.

In two models Z1 is in the highest level. He was the vice chief designer and had more knowledge in both knowledge areas. The approving task was performed by Z1. K1 was the leader of the Equipment Department and K2 was the leader of the General Department. They both performed the censoring tasks. B11 was the leader of the Ground Equipment Unit which included B12, B13, B14, B15 and B16. He performed the exam tasks. B12, B13, B14, B15 and B16 only originated the process. From the structure of these two models, it can be concluded that the knowledge maintenance organization model was made mainly according to the formal organization model.

In comparison to the predefined knowledge maintenance organization model, it is obvious that in the actual organization model, B68 is isolated from other users. Since there were no handover of knowledge between B68 and other uses, we could not get his real knowledge area and knowledge level.

The other difference is the level of B32. B32 is in the same knowledge area and at the same level with B34, B35 and B36 as shown in Figure6. However, he is at the same level with K1 and is involved in the other two knowledge areas in the mined knowledge maintenance organization model as shown in Figure5. It indicates that he performed the same kind of knowledge maintenance tasks with the department leader K1 although he is an ordinary designer in the predefined organization model. It can be deduced that the task delegation occurred in the Equipment Department. Actually K1 previously delegated the censoring task to B32. K1 was involved in a project that the institute and the company in the north west of China were cooperated on. K1 was always on a business trip to the north west of China. He believed that B32 had enough knowledge and was qualified to censor the knowledge. He delegated the task to B32 when he was away and revoked the delegation when he came back. Although it may be unofficial it reflects the importance of B32 and it means that B32 has more knowledge than the other designers in the department.

The explanation for the actual organization model leads to the additional conclusion that the higher density of the Equipment Department was mainly caused by the participation of two users in the censoring task. In addition, the isolation of B68 was another cause for the lower density of the General Department.

With the above findings the institute is to take steps to encourage the active people such as K1 to make persistent efforts. B32 is chose as a candidate for the censoring task in the optimized knowledge maintenance organization model.
4. Conclusion

This paper focuses on the application of process mining to the knowledge maintenance. We focused on obtaining insights into the knowledge maintenance process by looking at the control-flow and organizational perspective. For these perspectives which are (1) process perspective and (2) organization perspective, we presented some initial results. We have shown that it is possible to derive understandable models giving insights into the process with existing techniques. The case study illustrating the practical application was used to show how the knowledge maintenance benefits from the process mining. As future work, the mining from more perspectives and the development of new mining methods are essential. Moreover, more logs will be used to get a deeper understanding of the knowledge maintenance process.

5. Acknowledgement

This research was supported by the National Natural Science Foundation of China under Grant 71101153 and New Recruitments of China University of Petroleum-Beijing (QD-2010-06).

6. Reference


