SEMANTIC APPROACH UTILIZING DATA MINING AND CASE-BASED REASONING FOR IT SUPPORT SERVICE

Niloofer Shanavas[1], Shimmi Asokan[2]

[1]M.Tech. Student, Department of Computer Science & Engineering, Rajagiri School of Engineering & Technology, Kochi, Kerala, India

[2]Assistant Professor, Department of Computer Science & Engineering, Rajagiri School of Engineering & Technology, Kochi, Kerala, India

Abstract

Information Technology (IT) plays a very important role in all organizations. IT executives are constantly faced with problems that are difficult to tackle. Failure in IT service can interrupt the functioning of an organization. Case-Based Reasoning (CBR) is a problem solving methodology where experience in the form of past cases can be used to solve problems, thereby assisting the automation of problem solving and experience management. Furthermore, the performance, quality and efficiency of CBR systems can be enhanced through data mining. In order to support the IT team for faster and efficient problem resolution, a case-based reasoning approach integrated with data mining techniques could be utilized. In this paper, the study done on various CBR systems and data mining techniques for problem and experience management is explained. A system is proposed for IT experience and problem management with semantic retrieval in order to increase the efficiency and quality of the IT support service.

Keywords: Case-based reasoning, Data Mining, Experience management, IT problem management, IT support.

---***---

1. INTRODUCTION

Today, Information Technology is vital for most of the organizations and an important element for their functioning. Every organization has to invest in Information Technology to compete and succeed. Gathering and sharing IT knowledge can increase the organizational intelligence, allowing organizations to compete much better [1]. IT is strongly embedded in the business and has become an integral part of business management. Decisions on IT matters have brought about far reaching consequences for the business. IT executives are burdened with questions that are not easy to tackle. The IT team can resolve issues quickly and effectively if they have the information that can support them in their problem resolution. Collection of relevant data, analyzing it and proper interpretation is needed for obtaining this information [2].

The increasing significance of Information Technology resources and associated services that supports the organization´s business units have made Information Technology Service Management (ITSM) of vital importance for an organization [3]. Information Technology Infrastructure Library (ITIL), a methodology used in large organizations for IT management, is a widely accepted approach for IT Service Management across the world. ITIL provides a practical, no nonsense framework for identifying, planning, delivering and supporting IT services in the business [4]. ITIL framework emphasizes on Problem Management to reduce the number and severity of incidents, thereby reducing potential problems to the business/organization. Problem management consists of reactive and proactive aspects. Solving a problem when one or more incidents arise is the reactive problem management [5]. Identifying and solving potential causes of error before any incidents occur is the proactive problem management. Reusing IT knowledge and problem resolution history is of high value to reach these problem management goals. This helps us to detect the root causes of IT problems in a more effective way and help bring down the associated costs. In [7], case-based reasoning is defined as a reasoning model that assimilates problem solving, understanding, and learning, and integrates all of them with memory processes. CBR functions with the help of case-based knowledge. In [7], a case is defined as a contextualized piece of knowledge called experience that is required to attain the goals of the system. Comparisons between the new and old situations are used in a case-based system. This method of case-based reasoning depends on the assumption that "similar problems have similar solutions". This assumption that is referred to as the CBR hypothesis is the managing principle essential for most CBR systems [8]. Since case-based reasoning resolves new issues by using resolutions used to solve past issues, it is extensively used to develop support systems for IT management. Since it is possible that past issues can arise again in a similar fashion, CBR systems are very useful for problem management. Thus CBR automates the process of describing problems and defining solutions to these problems [9].

The CBR architecture consists of four prime components as shown in Fig. 1 [7]. In CBR system, similar previous cases are retrieved to solve a new problem. The retrieved cases suggest a solution for the problem. If necessary, the suggested solution can be revised for solving the new problem; this then becomes the confirmed solution. The new solution is retained as part of a new case and is called the learned case. This is the core process of the CBR cycle.
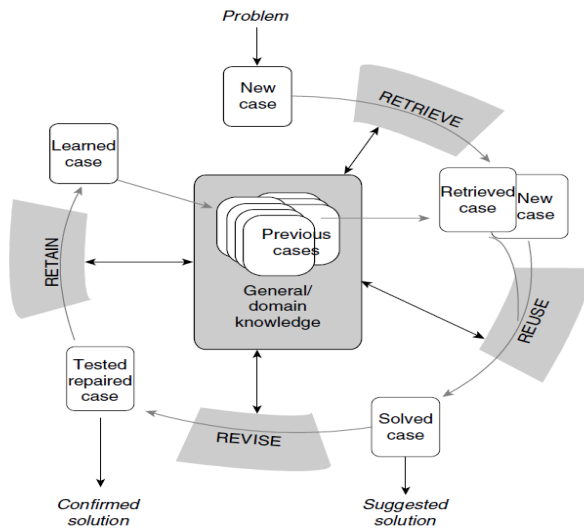

Fig 1 CBR cycle

With the help of data mining technology for feature selection and case retrieval, the quality of CBR systems can be improved [10]. A huge case consisting of a large number of features can be reduced to a feature vector with only a few principal features through feature selection. In a case base, the cases can be clustered into sub-case bases at different levels that make case management simple, case retrieval more efficient and accurate, thereby increasing the efficiency of CBR systems. Data Mining is an efficient and significant tool for knowledge discovery from a large database. New approaches to problem solving are obtained with the help of data mining for decision support by discovering patterns and relationships hidden in data.

2. IT PROBLEM MANAGEMENT

IT Problem management is one of the components in the ITIL support area. Problem management helps in reducing the number of incidents and potential problems in an organization [5]. Problem management minimizes problems and incidents due to errors within the IT infrastructure. Problems are assigned a priority level with the highest priority given to a problem that can interrupt critical IT services. Prioritization helps the IT support team to identify the impact and urgency associated with problems. If problems are not detected and solved, it can have a serious impact on the availability of IT services. An important task of problem management is to

document the information about the incident in order to support all problem management activities. Problem management has many advantages. It improves IT service quality, reduces the number of incidents, obtains permanent solution, improves organizational learning and increases fix rate [5]. Reliable IT services are very necessary for an organization. Incident reductions minimize the disruptions in the organization. Permanent solution for problems gradually reduces the number of problems. Problem management depends on learning from previous solved cases. Service desk can easily solve problems that are solved earlier and documented. An effective system that aids the IT executive in solving problems is essential for the success of problem management. In [6], a solution is proposed to find the root cause of problems from previous diagnoses of IT failures. Through interaction with IT operators, it finds the root cause of problems. Hence, importance is given to the reuse of knowledge and IT operator's experience. Earlier solutions were static, making it impossible to reuse knowledge to solve similar failures. Failures have to be corrected without any delay to maintain high quality service. The steps in problem resolution are observing service alerts, raising trouble tickets, assigning tickets, identifying root cause, planning resolution, implementing resolution and notification/clean up [15]. The major troubleshooting effort is spent on identifying the root cause and planning resolution. The troubleshooting effort needs to be reduced to save the time and number of resources required for problem resolution.

3. EXPERIENCE MANAGEMENT

In this busy and competitive world, one area that has gained prominence is experience management. Experience is required for complex problem solving. Experiences are shared to obtain the experiences of different people and avoid the kind of mistakes that other people have done. The problem need not be solved from scratch by reusing experience. Experience management is a particular kind of knowledge management [12]. A problem solver obtains specific valuable knowledge called experience while problem solving. The specific activities of experience management are to collect, model, store, reuse, evaluate and maintain experience. The experience base is a collection of experience items. In case-based reasoning, experience item is known as the case and case base is the other term for experience base. Approaches towards experience management include Experience Factory and Lessons Learned System. An approach called Experience Factory, particularly tailored to software engineering, captures, documents, stores and disseminates experience [11]. These experiences are stored in an experience base that forms organizational memory. Hence, the experience factory is an organizational framework that supports reuse and learning. Another system that implements knowledge management approach to collect, store, disseminate and reuse experiential knowledge is the lessons learned system [23]. CBR methodology that obtains solution of a new problem from

similar problems is applied as a foundation for organizational systems that manage experience [17].

4. CASE-BASED REASONING (CBR) SYSTEMS

CBR systems make use of the specific knowledge of past experiences (cases). A similar past case is found for a new problem and reused to obtain the solution for the new problem. In [13], it is mentioned that CBR is also an approach to incremental learning since a new experience obtained by solving a problem is made available for future problems. In [14], a system called DUMBO is presented that integrates case-based reasoning methodology in traditional trouble ticket system architecture that extends its functionalities by diagnosing new problems using previous network faults. DUMBO performs continuous learning, i.e. the knowledge base is gradually increased by the integration of new solved cases. It uses production rules in the reasoning process. Other systems that use cased-based reasoning in association with trouble ticket systems are MASTER and CRITTER [14]. A positive feature of CRITTER is that its adaptation strategy is automated and that of MASTER is its friendly user interface. The advantages of DUMBO compared to these systems are its simple structure, adds new relevant features in the learning and allows users interaction [14]. The disadvantage of DUMBO system is that the production rules for case matching are manually generated.

In [15], the information retrieval component of system is described that retrieves relevant articles from a collection of previously solved problems and their associated solutions. This system assists the system administrators in problem resolution by identifying similar problems and retrieving the steps taken to resolve those problems. A term vector is used to represent an article. Taxonomy is used to define the document representation space and it gives a source of keywords for the initial case description.

In [16], a web-based dynamic knowledge base system called NetPal is described. NetPal supports network administrators to troubleshoot tasks, recall and store experience, and identify new failure case and hence reduces time and resources required. It supports experience management and assists system administrators to recall by suggesting past experience cases. The system represents various domains like knowledge management, information retrieval, machine learning and network management. The vector space model is used to represent the cases for information retrieval and cosine angle distance is used to measure the similarity between vectors. Log Data Acquisition Engine collects and preprocesses log data from multiple services that are aggregated in a central database and indexed for retrieval. One drawback is that experience cannot be automatically collected for the case database.

5. DATA MINING FOR IT EXPERIENCE MANAGEMENT SYSTEMS

Data mining, also known as knowledge discovery in databases (KDD), is an emerging field. It helps to analyze, understand and visualize the vast amounts of stored data gathered from several business applications and help organizations to make better decision to stay competitive in the industry [21]. The introduction of data mining to information systems improves the quality of knowledge discovery process and decision process [22]. The techniques of data mining like clustering, classification, association rules and generalization could be used to optimize and improve Experience Management Systems [11]. Clustering is used to group similar experiences. Classification is used for the categorization of experiences into classes. Association rules are used to find relationships between data in a large database and generalization is used for data summarization. These data mining techniques are used in analyzing, generalizing, processing the experiences and building the experience base [11]. Frequently asked questions (FAQ) pages assist IT executives in responding to inquiries. In [18], a system is proposed that help a service company's operators respond to inquiries and construct FAQ. The system has four components: FAQ filtering, FAQ matching, replying support and FAQ building. The FAQ filtering part performs a comparison of inquiries with the questions in the FAQ to determine whether they are related. The FAQ matching part automatically replies with a question and answer set in FAQ if the inquiry is related to a question in the FAQ. The inquiries that are not related to any question in the FAQ are forwarded to the operators. The replying support component displays a reply form to operators based on inquiry and replying records from the past. In the FAQ building part, operators are displayed a FAQ that are constructed from past inquiries. Clustering technique is used in the FAQ building part to group similar inquiries and answer sets so that it becomes easier for the operators to analyze the records that can save their time. In [19], it is mentioned that there could be large amount of cost saving if the IT support team reduces the problem troubleshooting time. The proposed solution in [19] is an information retrieval approach combined with data mining techniques to extract information automatically from public resources such as FAQs and forums to generate a knowledge base for dynamic IT management support. This is especially beneficial for small to medium sized organizations that do not have many experienced IT staff. Crawlers are employed to collect experience data (problem/solution pairs) from publicly available resources such as FAQs, forums and web sites that are useful for the IT management team. The data is preprocessed by punctuation removal, stop word removal and stemming; each data instance is represented by a term vector using its term frequency/inverse document frequency (TF/IDF). Then, clustering techniques are used to group similar problem/solution pairs for quick retrieval. The clustering algorithm used is an improved version of that used in [18] as it reduces its computational cost. In [20], the

proposed system in [19] is improved by using the clustering algorithm with Multi objective genetic algorithm (MOGA), an optimization technique that automates the process of choosing clustering parameters.

6. PROPOSED SYSTEM

The studies conducted on data mining and case-based reasoning for IT support service, reveals that keyword-based retrieval has many limitations compared to concept-based retrieval/semantic retrieval. Synonymy (multiple words having similar meaning) and polysemy (words having more than one meaning) are the main causes for inaccurate retrieval in keyword-based systems since this results in mismatches between the vocabulary used by users and that used in the documents. An effective approach towards concept-based retrieval is to use Latent Semantic Indexing (LSI) / Latent Semantic Analysis (LSA), a soft clustering algorithm and a dimensionality reduction technique. Latent Semantic Indexing [24] is a statistical information retrieval method that retrieves text based on matching of concepts and not that of keywords. We propose a system with semantic retrieval as shown in Fig. 2 for IT problem and experience management that increases the efficiency and quality of the IT support service.
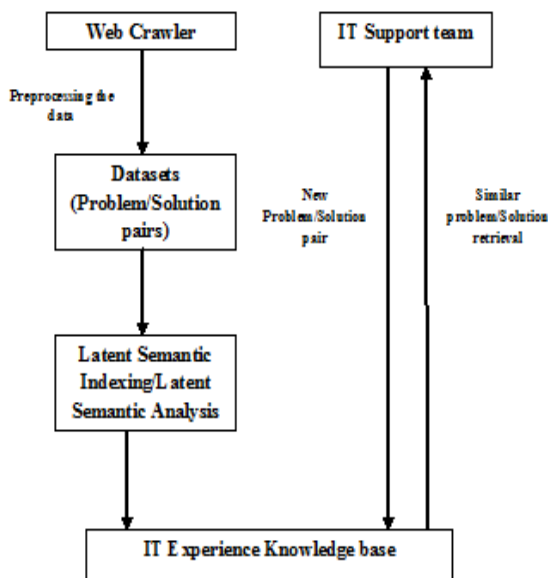

Fig 2: IT Experience and Problem Management System Design

The web crawler extracts IT experience data or solved IT problems from public resources. The extracted information is preprocessed to obtain the dataset containing problem/solution pairs. Latent semantic analysis is done on the documents in order to transform it into lower dimensional latent space. This brings out the hidden structure of these documents and represents the documents in a semantic way. The IT team can semantically retrieve documents (problem/solution pairs) corresponding to their problem. They can also add new problem/solution pairs to the knowledge base which results in a dynamically incrementing IT experience knowledge base. The proposed system guarantees accurate and efficient retrieval of similar problem/solution pairs from the IT experience knowledge base.

7. CONCLUSIONS & FUTURE WORK

Information Technology support is an area that requires vast amount of knowledge. Trouble shooting time could be reduced if the IT experience gained could be accumulated in an experience base. Hence, IT experience management is valuable. Case-based reasoning approach which is guided by the principle that similar problems have similar solutions is widely used to develop support systems for IT management that solves new problems by using or adapting solutions used to solve old problems. This paper explains how CBR systems integrated with data mining techniques can automate IT experience and problem management and proposes an efficient system for IT support service. Focusing more on approaches for semantic processing can continue research in this area.

REFERENCES

[1] Paul Klint, Chris Verhoef, "Enabling the creation of knowledge about software assets", Data and Knowledge Engineering e41 (23), pp. 141-158, 2002.

[2] L.M. Kwiatkowski, C. Verhoef, "Recovering management information from source code", Science of Computer Programming, Volume 78, Issue 9, pp. 1368-1406, 2013.

[3] Jacques P. Sauve, Rodrigo A. Santos, Rodrigo R. Almeida, J. Antao B. Moura, "On the Risk Exposure and Priority Determination of Changes in IT Service Management", 18th IFIP/IEEE International Workshop on Distributed Systems: Operations and Management (DSOM 2007), pp. 147-158, 2007.

[4] http://www.itil-officialsite.com/

[5] http://www.ucisa.ac.uk

[6] Ricardo L. dos Santos, Juliano A. Wickboldt, Roben C. Lunardi, Bruno L. Dalmazo, Lisandro Z. Granville, Luciano P. Gaspary, Claudio Bartolini, Marianne Hickeyt, "A solution for identifying the root cause of problems in IT change management", Integrated Network Management (IM), 2011 IFIP/IEEE International Symposium, pp. 586-593, 2011.

[7] Shiu S., Pal S. K., "Foundations of soft case-based reasoning", John Wiley & Sons, 2004.

[8] Hullermeier E., "Toward a probabilistic formalization of case-based Inference", Proceedings of the 16th international joint conference on artificial intelligence (IJCAI99), Stockholm, Sweden, pp. 248-253, 1999.

[9]   T. Olsson, P. Funk, "Case-based reasoning combined with statistics for diagnostics and prognosis", 25th International Congress on Condition Monitoring and Diagnostic Engineering, Journal of Physics: Conference Series, 2012.

[10] Yuan Guo, Jie Hu, Yinghong Peng, "Research on CBR system based on data mining", Applied Soft Computing Volume 11, Issue 8, pp. 5006-5014, 2011.

[11] Klaus-Dieter Althoff, Bjorn Decker, Susanne Hartkopf, Andreas Jedlitschka, Markus Nick, Jorg Rech, "Experience Management: The Fraunhofer IESE Experience Factory", Proc. Industrial Conference Data Mining, 2001.

[12] Bergmann Ralph, "Experience Management: Foundations, Development Methodology, and Internet-Based Applications", Springer, 2002.

[13] Aamodt, E. Plaza, "Case-Based Reasoning: Foundational Issues, Methodological Variations and System Approaches", AI Communications, IOS Press, Vol. 7: 1, pp. 39-59, 1994.

[14] Melchiors, L. M. R. Tarouco, "Troubleshooting network faults using past experience", Network Operations and Management Symposium, pp. 549-562, 2000

[15] George A., Makanju A., Zincir-Heywoo A.N., Milios E.E., "Information Retrieval in Network Administration", Communication Networks and Services Research Conference, pp. 561-568, May 2008.

[16] Ashley George, Adetokunbo Makanju, Evangelos Milios, Nur Zincir-Heywood, Markus Latzel, Sotirios Stergiopoulos, "NetPal: a dynamic network administration knowledge base", CASCON '08: Proceedings of the 2008 conference of the center for advanced studies on collaborative research, 2008.

[17] Klaus-Dieter Althoff, Rosina O. Weber, "Knowledge management in case-based reasoning", The Knowledge Engineering Review, Vol. 20:3, Cambridge University Press, pp. 305-310, 2006.

[18] Iwai K., Iida K., Akiyoshi M., Komoda N. "A help desk support system with filtering and reusing e-mails", Industrial Informatics (INDIN), 2010 8th IEEE International Conference, pp. 321-325, 2010.

[19] Can Bozdogan, Nur Zincir-Heywood, "Data Mining for Supporting IT Management", Network Operations and Management Symposium (NOMS), IEEE, pp. 1378-1385, 2012.

[20] Bozdogan C., Zincir-Heywood A.N., Gokcen Y., "Automatic optimization for a clustering based approach to support IT management", Integrated Network Management (IM 2013), 2013 IFIP/IEEE International Symposium, pp. 1233-1236, 2013.

[21] S.C. Hui, G. Jha, "Data mining for customer service support", Information & Management, Volume 38, Issue 1, pp. 113, October 2000.

[22] Rok Rupnik, Matjaz Kukar, "Data Mining and Decision Support: An Integrative Approach", Decision Support Systems, Book edited by Chiang S. Jao, 2010.

[23] R. Weber, D.W Aha, I. Becerra-Fernandez, "Intelligent lessons learned systems", Expert Systems with Applications, Volume 20, Issue 1, pp.17-34, 2001.

[24] Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, Richard Harshman, "Indexing by latent semantic analysis", Journal of the American Society for Information Science, pp. 391–407, 1990.