Analysis of Requirements Engineering Techniques Using Clustering

[1]Li Jiang [2]Armin Eberlein

[1] Dept. of Computer Science and Applied Statistics

University of New Brunswick, Canada

[2] Computer Engineering Department

American University of Sharjah, UAE

Abstract: The selection of RE techniques for a project is usually based on personal preference or existing company practice rather than on characteristics of RE techniques and the project at hand. Moreover, research has shown that there are a lot of very useful RE techniques that are not widely used. The few approaches currently available for the selection of RE techniques provide only little guidance for the actual selection process. We believe that the thorough understanding and evaluation of RE techniques in the context of an application domain and a specific project is of great importance. This paper describes research that analyzes RE techniques using a clustering method. An industrial case study that integrated the results of the clustering into the RE technique selection process illustrated the valuable help provided by the clustering.

Keywords: Requirements engineering, technique evaluation, clustering, decision support.

1 Introduction

Requirements engineering (RE) is an essential part of the software development process. It plays an important role in ensuring the overall quality of a software product [1-3]. Currently, there are numerous techniques that address different aspects of the RE process and system development [4] and that can be applied to various types of projects. However, the application of RE techniques in the context of a specific software project is never trivial due to the inherently uncertain nature of RE. This is one of the big challenges faced by the RE community. So far, only very limited work has been done that provides help for the analysis of RE techniques, such as in [5-9]. However, all the current research into technique selection does not provide sufficient analysis of RE techniques, therefore, the logical link between a software project and the rationale for the use of certain RE techniques in the software project is either missing or very week.

The objective of this research is to provide help for the selection of the most suitable RE techniques by explicitly considering the characteristics of RE techniques and the software project under development. This would meet the critical need of industry for advice on how and when to use certain RE approaches [10-11].

The rest of the paper is organized as follows: Section 2 gives a more detailed discussion of the techniques analysis. The conclusion and future work is presented in Section 3.

2 Analysis of RE Techniques

In our earlier research, we have identified a subset of 46 RE techniques (see Table 1) [4, 12-13]. The techniques selected are representative RE techniques that cover all phases of the RE process. Additionally, they are also widely known in industry, have a well-defined scope and are well-documented. Newly identified techniques will be included in our future research.

RE techniques analysis includes the following procedures:
1) Identification of the attributes of RE techniques.
2) Evaluation of the RE techniques currently documented in our research. Each technique is rated against 31 attributes by researchers and experts from both industry and academia.
3) Detailed analysis of the techniques using a clustering method.

The following subsection briefly summarizes the basic procedures of the analysis and its results:

2.1 Identification of the attributes of RE techniques

Classification of RE techniques requires the definition of attributes that characterize these techniques. The 46 RE techniques analyzed and compared during this research are listed in Table 1 together with the major stages in which these techniques can be applied.

Based on our analysis, 31 attributes (see Table 2) for RE techniques were defined [12]. The first column in Table 2 contains the categories of the attributes which correspond to the four stages of the RE process. The third column lists the actual attributes. Each attribute is defined with a list of criteria to ensure its measurability. An ordinal scale is used for all attributes, i.e. the attribute values are set as none (or “not relevant”), very low, low, medium, high and very high. The interested reader can refer to [12] for more details.

As can be seen in Table 2, the attributes in the schema provide a means to measure different facets of an RE technique. The attributes can, therefore, be divided into two categories:

- Attributes that describe the ability of a technique (attributes 1 to 28): The higher the value of an attribute, the more suitable is the technique for addressing this attribute.
- Attributes that describe economic factors (attribute 29 to 31): The higher the value of an attribute, the higher the cost of using the technique.

This classification is essential for the calculation of the abilities and cost of the RE techniques. Details of the usage of the schema will be discussed in Section 2.2.

Table 1 Summary of analyzed RE techniques

ID	Technique Name	Most Common Area of Application in the RE Process
1	Brain Storming and Idea Reduction	Elicitation
2	Designer as Apprentice	Elicitation
3	Document Mining (Observation)	Elicitation
4	Ethnography	Elicitation
5	Focus Group	Elicitation
6	Interview	Elicitation
7	Contextual Inquiry	Elicitation
8	Laddering	Elicitation
9	Viewpoint-Based Elicitation	Elicitation (later stage)
10	Exploratory Prototypes (Throw-Away Prototype)	Elicitation, Analysis and Negotiation, Verification and Validation
11	Evolutionary Prototypes	Elicitation, Analysis and Negotiation, Verification and Validation
12	Viewpoint-Based Analysis	Analysis and Negotiation
13	Repertory Grids	Requirements Elicitation
14	Scenario Approach	Requirements Elicitation (later stage) , Requirements Analysis and Negotiation, Documentation, Verification and Validation
15	JAD	Elicitation
16	Soft Systems Methodology (SSM)	Elicitation
17	Goal-Oriented Analysis	Analysis and Negotiation
18	Viewpoint-Based Documentation	Documentation
19	Future Workshop	Elicitation
20	Representation Modeling	Analysis and Negotiation, Elicitation
21	Functional Decomposition	Analysis and Negotiation
22	Decision Tables	Analysis and Negotiation, Documentation, Verification
23	State Machine	Analysis and Negotiation, Documentation, Verification
24	State Charts (also known as State Diagrams)	Requirements modeling, Documentation, Verification
25	Petri-nets	Analysis and Negotiation, Documentation, Verification
26	Structured Analysis (SA)	Analysis and Negotiation, Documentation, Verification
27	Real Time Structured Analysis	Analysis and Negotiation, Documentation, Verification
28	Object-Oriented Analysis (OOA)	Analysis and Negotiation, Documentation, Verification
29	Problem Frame Oriented Analysis	Analysis and Negotiation, Documentation, Verification
30	Goal-Oriented Verification and Validation	Verification and Validation
31	ERD (Entity Relationship Diagram)	Documentation
32	AHP	Requirements Prioritization
33	Card Sorting	Requirements Prioritization
34	SQFD (Software QFD)	Analysis and Negotiation and Elicitation
35	Fault Tree Analysis	Analysis and Negotiation and Elicitation
36	Structured Natural Language Specification	Requirements Documentation
37	Viewpoint-Based Verification and Validation	Verification and Validation
38	Unified Modeling Language (UML)	Documentation, Analysis and Negotiation, Verification
39	Z	Documentation, Analysis, Verification
40	LOTOS	Documentation, Analysis, Validation
41	SDL	Documentation, Analysis, Validation
42	XP (Extreme Programming)	Elicitation, Analysis and Negotiation, Documentation, Validation
43	Formal Requirements Inspection	Requirements Verification and Validation
44	Requirements Testing	Requirements Verification and Validation
45	Requirements Checklists	Requirements Verification and Validation
46	Utility Test	Requirements Verification and Validation

2.2 Techniques Evaluations

In this research, 46 techniques were assessed against each attribute in the schema as described in Table 2. Detailed assessment information can be found in [12].

Examples of the assessment are shown in Table 2. Columns 4 to 7 contain the normalized results of the assessment of 4 techniques. For example, the communication ability of interviewing techniques is assessed as “Very High”. The normalized value for “Very High” is 1, i.e., the entry for that column is 1. This data set is the foundation for the further analysis of RE techniques. A complete list of the assessment results can be found in [12].

Table 2 An attribute schema for RE techniques and their assessment

Categories	No.	Attributes of the Techniques	Interview	JAD	State Charts (also known as State Diagrams)	XP Techniques
Elicitation:	1	Ability to facilitate communication	1	1	0	1
	2	Ability to understand social issues	0.6	1	0	0.4
	3	Ability to get domain knowledge	0.6	0.6	0	0.4
	4	Ability to get implicit knowledge	0.2	0.2	0	0
	5	Ability to identify stakeholders	1	1	0	0.4
	6	Ability to identify non-functional requirements	1	0.8	0	0
	7	Ability to identify viewpoints	0.8	1	0	0
Analysis & Negotiation:	8	Ability to model and understand requirements (both general and domain specific requirements)	0	0	1	0.6
	9	Ability to understand the notations used in analysis	0	0	0.8	0.8
	10	Ability to analyze non-functional requirements	0	0	0	0
	11	Ability to facilitate negotiation with customers	0	0	0.4	1
	12	Ability to prioritize requirements	0	0	0	1
	13	Ability to identify accessibility of the system	0	0	0.6	0.2
	14	Ability to model interface requirements	0	0	0.6	0.8
	15	Ability to identify reusable requirements and support requirements reuse	0	0	0	0
Documentation & Notation:	16	Ability to represent requirements (Expressibility)	0	0	1	0.2
	17	Capability for requirements verification	0	0	0.8	0.4
	18	Completeness of the semantics of the notation	0	0	0.6	0
	19	Ability to write unambiguous and precise requirements by using the notation	0	0	0.8	0.2
	20	Ability to write complete requirements	0	0	0.6	0.2
	21	Capability for requirements management	0	0	0	0.4
	22	Modularity	0	0	0	0
	23	Implementability (Executability)	0	0	0	0
Verification & Validation	24	Ability to identify ambiguous requirements	0	0	0.6	0.6
	25	Ability to identify interactions (ambiguous, inconsistency, conflict)	0	0	0.2	0.2
	26	Ability to identify incomplete requirements	0.2	0	0	0
Other Aspects	27	Ability to support COTS-based RE process	0	0	0	0
	28	Maturity of supporting tool	0	0.4	0.6	0.4
	29	Learning curve (Introduction cost)	0.2	0.6	0.6	0.4
	30	Application cost	0.4	0.6	0.6	0.2
	31	Complexity of techniques	0.2	0.2	0.4	0.2

2.3 Techniques Clustering

Clustering has been used extensively as a data analysis technique in various domains, such as medical data analysis, data mining, and market analysis. Cluster analysis organizes data by abstracting the underlying structure either as a group of individuals or as a hierarchy of groups [14]. This means that clustering techniques allow objects with similar attributes to be organized into groups.

Table 3 Clustering setting and performance

Setting / Performance using the setting	Number of clusters=4		Number of clusters=5		Number of clusters=6		Number of clusters=7		Number of clusters=8		Number of clusters=9		Number of clusters=10		Number of clusters=11		Number of clusters=12	
	All weights are 1 (W_i=1)	Various weights	All weights are 1 (W_i=1)	Various weights	All weights are 1 (W_i=1)	Various weights	All weights are 1 (W_i=1)	Various weights	All weights are 1 (W_i=1)	Various weights	All weights are 1 (W_i=1)	Various weights	All weights are 1 (W_i=1)	Various weights	All weights are 1 (W_i=1)	Various weights	All weights are 1 (W_i=1)	Various weights
Value of the cost function	27.12	12.96	25.43	10.65	21.78	6.18	18.81	5.11	17.55	3.60	17.92	3.64	18.12	5.50	19.55	5.87	19.38	6.81
Clustering effects	Basic	Basic	Basic	Basic	Better	Better	Better	Better	Best	Best	Best	Best	Better	Better	Better	Better	Better	Better

Notes: 1. The values of the cost function are calculated based on formula (C-1)

2. In the row containing the Clustering effects, "Basic" indicates that the result of the clustering gives a rough indication of the high level class to which a RE technique belongs. However, the techniques in each cluster do not have very similar characteristics. "Better" indicates that the result of the clustering provides better classification results. "Best" indicates that the result of the clustering provides a fine-grained classification of the RE techniques. The techniques in each cluster exhibit very similar characteristics as can be verified by RE experts.

3. The "weight" refers to the weight of each attribute of the techniques; "various weights" indicates that the attributes were assigned different weights based on the characteristics of the project.

There are a number of clustering methods, such as K-Means, Hierarchy Clustering and Fuzzy Clustering. The Fuzzy Clustering method is used in this research to analyze the similarity between RE techniques since the data derived in this research is fuzzy in nature.

The basic principle of the Fuzzy Clustering algorithm is to partition *n* objects into *p* clusters by minimizing the following cost function [15, 16]:

$$\text{Cost} = \sum_{i=1}^{p}\sum_{j=1}^{n} u_{ij}^{2} d_{ij}^{2} \qquad \text{(C-1)}$$

where

$d_{ij} = \| X_j - m_i \|$ is the distance between each object and the cluster centroid m_i, j=1,…, n; n is the number of objects, i=1, …, p; p is the number of clusters.

m_i is a vector representing the centroid of cluster i,

u_{ij} is the degree of membership of object j in cluster i.

The cost function is minimal if

$$m_i = \frac{\sum_{j=1}^{n} u_{ij}^{2} X_j}{\sum_{j=1}^{n} u_{ij}^{2}} \quad \text{and} \quad u_{ij} = \frac{1}{\sum_{k=1}^{p}\left(\frac{d_{ij}}{d_{kj}}\right)^{2}} \qquad \text{(C-2)}$$

46 RE techniques were clustered using 9 different settings as shown in the first two rows of Table 3. This was done in order to identify the optimum number of clusters that minimizes the cost function and ensures that the techniques in each cluster have similar characteristics. The results of the 9 different clustering trials can be seen in Table 3.

Based on the results of the calculations shown in Table 3, the following observations can be made:

- The higher the number of clusters the more fine-grained is the classification of the techniques. Further analysis shows that the classification into 8 or 9 clusters maximizes the cohesiveness of the techniques in each cluster and minimizes the cost function.
- Clustering provides a mechanism to group RE techniques according to their characteristics. Similar techniques are grouped into the same cluster thus providing a foundation for the analysis of the similarity of various techniques. It can therefore be used for the selection of RE techniques.
- Clustering also has its challenges. Some techniques were classified into an unsuitable cluster. Such misplacements are due to various reasons. Further research will focus on improving the Fuzzy Clustering algorithm and adapt it to the specific purpose of our research so that it will work more reliably.

Three concepts were developed to describe the relationship of techniques: functionally comparable techniques, functionally complementary techniques, and mutually exclusive techniques. The functionally comparable and functionally complementary techniques help the requirements engineers to identify those techniques that have the maximum ability and least cost. For instance, identification of functionally comparable techniques enables requirements engineers avoid using techniques that duplicate functionality thus minimizing the cost of the application of RE techniques. Additionally, using complementary techniques wisely ensures that the selected techniques have maximum ability to accomplish the mission of RE. A more detailed elaboration of these concepts can be found in [17].

3 Conclusion and Future Work

Previous research has shown that choosing proper RE techniques for system development contributes to the overall success of a project [4, 10]. Moreover, RE techniques selection has to be based on a thorough understanding of RE techniques and the relationships between them. The result of this research has shown that the clustering method is a helpful mechanism for analyzing RE techniques. Two case studies were conducted in industry where RE techniques clustering was used to offer help during RE techniques selection. The case studies indicate that clustering RE techniques was an effective means for the selection of the most suitable RE techniques [17-18].

Our future work includes further improvement of the performance of the Fuzzy Clustering Methods currently used, investigating more RE techniques, and applying clustering method to analyze these additional RE techniques. Since the RE technique clustering is only part of our overall research project called FRERE [17], the seamless integration of the clustering mechanism into the framework still has to be done.

References

[1] Boehm BW (1981) Software Engineering Economics, Prentice Hall

[2] Macaulay LA (1996) Requirements for Requirements Engineering Techniques. IEEE Proceedings of ICRE, pp. 157 – 164.

[3] Sommerville I, Sawyer P (1997) Requirements Engineering: A Good Practice Guide, John Wiley and Sons

[4] Jiang L, Eberlein A (2003) Requirements Engineering: A Review and A Proposal, Proceedings of the Third ASERC Workshop on Quantitative and Software Engineering, Feb. 17-18, 2003, Banff, Alberta, Canada.

[5] Maiden N, Rugg G (1996) ACRE: Selecting Methods for Requirements Acquisition. Software Engineering Journal, 11(3): 183-192, 1996.

[6] Hickey AM, Davis AM (2003) Requirements elicitation and elicitation technique selection: a model for two knowledge-intensive software development processes, Proceedings of the 36th Annual Hawaii International Conference on System Sciences, 2003., 6-9 Jan. 2003, Page(s): 96 -105

[7] Hickey AM, Davis AM (2003) Elicitation Technique Selection: How Do Experts Do It?, Proceedings of the 11th IEEE International Requirements Engineering Conference, 2003

[8] Bickerton JM, Siddiqi J (1993) The Classification of Requirements Engineering Methods, Proceedings of IEEE International Symposium on Requirements Engineering, 4-6 Jan 1993, pp. 182 - 186, San Diego, CA ,USA

[9] Kotonya G, Sommerville I (1998) RE, Processes and Techniques, John Wiley & Sons Ltd

[10] Jiang L, Eberlein A, Far BH (2003) Evaluating the Requirements Engineering Process Using Major Concerns, Technical Reports, August, 30, 2003, Dept. of Electrical and Engineering, University of Calgary, http://www.enel.ucalgary.ca/~jiangl/research/indexOfResearch.htm

[11] Glass RL (2004) Matching Methodology to Problem Domain, Communications of The ACM, May 2004, Vol. 47, No. 5, page 19-21

[12] Jiang L, Eberlein A, Far BH, Majid M (2003) An Innovative Requirement Engineering Technique Selection Model, Technical Reports, August, 30, 2003, Dept. of Electrical and Engineering, University of Calgary, http://www.enel.ucalgary.ca/~jiangl/research/indexOfResearch.htm

[13] Jiang L, Eberlein A, Far BH (2003) A RE process Template for Aiding RE process development, Technical Reports, August, 30, 2003, Dept. of Electrical and Engineering, University of Calgary, http://www.enel.ucalgary.ca/~jiangl/research/indexOfResearch.htm

[14] Jain AK, Dubes RC (1998) Algorithms for Clustering Data. Prentice-Hall

[15] Chiu S (1994) Fuzzy Model Identification Based on Cluster Estimation, Journal of Intelligent & Fuzzy Systems, Vol. 2, No. 3, Sep. 1994.

[16] Bezdek JC (1981) Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York

[17] Li Jiang, A Framework For Requirements Engineering Process Development, PhD Thesis, University of Calgary, Sep. 2005

[18] L. Jiang, A. Eberlein and B. H. Far, Combining Requirements Engineering Techniques –Theory and Case Study, ECBS 12th IEEE International Conference on the Engineering of Computer-Based Systems, Greenbelt, MD., USA, April 2005,