Volume 5, Issue 1, January 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Identifying Critical Human Factors that Impact Software Process in Medium Sized Enterprises Using Genetic Algorithm Based Association Rule Mining

A. Saranya
Research Scholar,
Madurai Kamaraj University,
Madurai, India

Dr. S. Kannan
Associate Professor,
Madurai Kamaraj University,
Madurai, India

Abstract: With the increasing number of medium sized software organizations, production of quality software by such organizations has become a huge challenge for the stakeholders of such organizations. It is a well appreciated fact that a quality focus on the process is likely to yield quality products. While addressing the issue of Software Process Improvement in such organizations, the role played by the human factors seems to be the most ignored of all. The present paper is aimed at uncovering some of the interesting associations between these human factors and the quality of the process in medium sized organizations by harnessing the power of data mining techniques

Keywords: IASRI, CMMI,

I. INTRODUCTION

The American Heritage dictionary defines quality as "an attribute or characteristic of something" [1]. Often these attributes are measurable – like cost, color, length. But given the nature of software, it is difficult to characterize software quality. Many innovations in the software engineering domain, notably, the CMMI [2] have endeavored to improve the process in an attempt to inculcate quality.

Medium sized software organizations face a large number of challenges, the most important being the paucity of resources that can be allocated to quality. It is necessary to evolve separate software engineering principles for such organizations in their attempts of software process improvement.

II. HUMAN FACTORS

While addressing the issue of Software Process Improvement in medium sized software organizations, the role played by the human factors is one of the most critical aspects to be considered. Human factors engineering pertains to the application of scientific knowledge concerning human behavior to the design of jobs and products. While a lot of works identify the human factors involved in software processes, little has been done in the application of scientific methodologies to draw meaningful and valid conclusions pertaining to human factors.

White addresses the people factors that have an impact on productivity of software [3]. White suggests that it is important to have the right people, people who are knowledgeable, skilled and satisfied, project manager who works well with people to ensure quality.

Since software is developed by people, human factors have a strong impact on the success of software development projects [4]. Much of the software engineering research in the last decade has been technical ignoring the people aspect. Modern software process models especially those in the agile paradigm, place a significant importance to human factors.

III. ASSOCIATION RULE MINING

Data Mining is generally considered as the process of discovering hidden, non-trivial and previously unknown patterns from a large collection of data. Association rule mining is an important component of data mining. According to the Indian Agricultural Statistical Research Institute (IASRI), association rule mining is perhaps the most widely studied model by the data mining community [5]. Potential examples of the utility of the mined association rules abound in almost every discipline. Organizations, particularly, in the retail segment are interested to use association rule mining to discover customer buying patterns that can aid them in major decision making regarding which products can be promoted together. Medical domain is no exception. Association rules can help doctors find the factors most likely to cause a particular disease. This is especially significant for non-communicable diseases for which the importance of various contributing factors is relatively unknown. Association rule mining can also be applied to agricultural databases to survey data from agricultural research.

The basic objective of association rule mining is to find all co-occurrence relationships called associations. The classic application of association rule mining is the market basket analysis which aims to discover how items purchased by customers are associated. An association rule is of the form $X \rightarrow Y$ where X and Y are collections of attributes whose intersection is null. For example, every customer who purchased a computer (X) also purchased a printer(Y). X is called the antecedent and Y is called the consequent. Since the number of possible association rules can be huge, often the interest is on those rules which satisfy some constraints. The most common of these constraints include support and confidence.

Formally, let I={i1, i2, …, im} be a set of items. Let T={t1, t2, …, tn} be a set of transactions where each transaction ti is a s et of items. An association rule is of the form X → Y, where $X \subset I, Y \subset I$ and $X \cap Y = \phi$. X (and Y) is a set of items called an itemset. The support of a rule X → Y is the percentage of transactions in T that contains X U Y and can be stated as Pr(XUY) which is the estimate of probability. If n is the number of transactions then the support of the rule XUY is given as (XUY).count/n. Confidence of a rule X → Y is the percentage of transactions in T that contain X also contain Y. it can be stated as Pr(Y|X) – the conditional probability.

Many algorithms exist for discovering Association Rules. Apriori and FP growth are popular representatives in this category.

IV. GENETIC ALGORITHMS

Genetic Algorithms belong to the class of evolutionary computation and attempt to mimic the natural process of evolution to uncover solutions to problems. The basic idea behind them were presented in the previous chapter. The pseudo-code is sketched here for completeness.

1. Initialize the population by generating an initial set of chromosomes representing solutions
2. Calculate fitness of the individuals in the population based on some criteria
3. Select 2 individuals from the population at random based on the fitness so that the more fit individuals get more chance for selection
4. Perform cross over by selecting one or more cross points
5. Select chromosomes at random and Mutate them with some probability
6. If the terminating condition is met stop and display the best solution obtained. Else repeat steps 2 – 5 using the newly generated solutions

Genetic Algorithms have been extensively applied in various domains and are very promising in yielding solutions to complex problems involving a lot of uncertainty.

V. MINING ASSOCIATION RULES BETWEEN HUMAN FACTORS AND QUALITY OF SOFTWARE

The importance of human factors in software process improvement was elucidated in the previous sections. The research aims at deriving associations between the human factors that impact the software process quality in medium sized organizations.

An exhaustive list of all the human factors that impact the processes tends to be very large and lends itself impractical to be applied in any study. For the purpose of the research the following set of human factors is considered. The list is by no means complete and can be extended to accommodate many more.

Table 4.1 – List of Human Factors Considered

Human Factor	Description
Knowledge of Software Engineering Principles	The knowledge in the domain of software engineering
Experience in Software Process Improvement	The experience of the individual in participation in software process improvement endeavors
Previous Experience in software development	The experience in the domain of software development
Ability to work in a team	The ability of the individual to contribute in a team
Communication Skills	The ability of the individual to comprehend requirements and communicate problems faced
Motivation Level	The level of motivation of the individual
Commitment	The commitment of the individual toward the success of the project
Responsibility	The willingness of the individual to learn from past mistakes

5.1 Methodology

A project manager working in a local medium sized software organization and who has 12 years of experience in software development and 8 years in software project management is asked to rate the individuals in the organization that has successfully delivered more than 20 products. The rating is one on a five point scale from 1 to 5 with a 1 indicating the least and 5 indicating the maximum rating. For example, a rating of 0 for experience means the individual has no experience and a rating of 5 for motivation level means that the individual is motivated to the greatest extent. Despite the subjectivity inherent in such a rating, the experience of the manager is expected to increase the objectivity to a reasonable level and thus improve the quality and utility of the mined association rules. The information pertaining to the processes adopted for all the products delivered by the organizations is available for the study. The manager also rated the process followed in each of the projects. Here, a rating of 1 indicates a good process while a rating of 0 indicates unsatisfactory process.

5.2 Applying GA to mine association rules

To apply Genetic Algorithm to any problem the first issue that needs to be addressed is encoding of the solutions. Solutions have to be represented as chromosomes. When applying Genetic Algorithm to mine association rules there are basically 2 approaches for solution representation. In the Pittsburgh approach where each chromosome represents a set of rules and the Michigan approach where each chromosome represents a single rule. The research uses a modified Michigan approach proposed by Ghosh and Nath [6]. The basic idea is to associate 2 bits with each attribute. If these 2 bits are 00, then the following attribute appears in the antecedent part and if these 11, then the following attribute appears in the consequent part. The remaining 2 combinations – 01 and 10 imply the absence of the following attribute in the rule. Since, the position of various attributes in the chromosomes is fixed and the attributes are numeric, the values of the attributes can be encoded in their binary form. The values to be encoded are in the range of 0-3 for the ratings and 0-1 for quality attribute. 8 human factors would need 2 bits each and the quality attribute needs 1 bit. For all the 9 attributes 2 tag bits are required as described above. The total space requirement for association rules of arbitrary length turns out to be 8 * 2 + 1 + 9 * 2 = 35 bits. The fitness evaluation is done using an approach analogous to the one adopted by Wakabi-Waiswa and Baryamureeba [7]. The Mutation probability was set at 0.01. The number of generations is capped at 150.

5.3 Results and Discussion

The best performing individuals in the 150 generations are sorted according to their fitness values and from the resulting 150 rules, rules for which the consequent is some attribute other than the quality attribute are pruned. The top 5 association rules that were mined are shown in the table.

Table 2 - Top 5 Rules with the Process Attribute as the consequent (in terms of fitness)

Motivation Level=5 ^ Knowledge of Software Engineering Principles > 3 → Process=1
Previous Experience in software development > 3^ Ability to work in team > 4 → Process=1
Motivation < 2 ^ Commitment < 2 →Process=0
Experience in Software Process Improvement > 3 ^ Commitment > 3 → Process=1
Commitment>= 3 ^ Responsibility >= 3 → Process=1

The top 5 rules uncovered by tge research throw some light on the significant human factors that contribute to quality software processes in medium sized organizations. Motivation and Commitment play a significant part. If these are low, the process is likely to suffer. Motivation, commitment and responsibility seem to contribute greatly to good software processes. Knowledge of software engineering principles and previous experience can serve as added advantages but they by themselves do not necessarily lead to quality process.

5.4 Comparison with Apriori and FP Growth

To demonstrate the potential of GA, the same experiment was done with the Apriori Algorithm and the FP Growth algorithm. The following table gives a comparative statement about the performance of the algorithms in terms of execution time and average quality. In terms of time Apriori seemed to be the worst performer and this is expected given the time required for generating candidate item sets by the algorithm. Though the difference between GA and FP Growth in terms of time is very small, there is a considerable improvement in the quality of the rules generated. GA seems to generate more quality rules than FP Growth. This reinforces the opinion that as the complexity of the problem and uncertainty involved increases, trends shift in favor of GA.

Table 4.3 – Comparison of Apriori, FP Growth and GA

	Time (in secs)	Average Quality
Apriori	2.9	71.2%
FP Growth	1.7	74.6%
Genetic Algorithm	1.3	78.2%

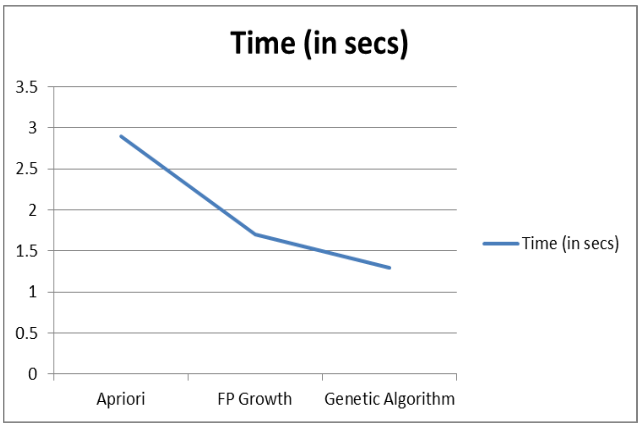

Figure 4.2 – Comparison in terms of time

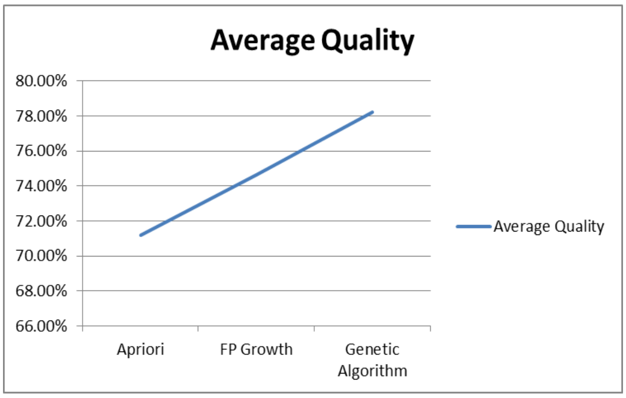

Figure 4.3 – Comparison of Apriori and GA in terms of average quality

VI. CONCLUSION

A genetic algorithm was propsoed to mine association rules to discover interesting asociations between human factors that contribute to software process quality in medium sized organizations. Human factors have a great impact on software processes and a careful identification of the significant human factors contributing to high quality processes is of utmost importance for project managers and organizations in their endeavor to improve their software processes. Several association rules are mined and the mined association rules indictate that motivation and commitment are two of the crucial human factorsin this regard. Knowledge of software engineering principles and previous experience can serve as added advantages but they by themselves do not necessarily lead to quality process. The research attemtps to throw light on the prospects of utilizing Association rule miniung and genetic algorithm in identifying the crucial human factors impacting processes. Given the strong correlation between human factors and quality this can be a starting step of an attempt that can lead to perceivable improveemnts in the quality of the software developed. Managers can reap great benefits by paying attention to the crucial human factors identified. In future, the list of human factors considered can be extended to include more and intresting associations discovered between them and the quality of software processes.

REFERENCES

[1] Pressman, Roger S, “Software Engineering – A Practitioner’s Approach”, Fifth Edition, 2001.
[2] “CMMI”, http://www.sei.cmu.edu/cmmi/
[3] K. S. White, Productivity And Quality, Jefferson Lab, Newport News, VA, USA
[4] John, Michael, Maurer, Frank, Tessem, Bjomar, Human And Social Factors of Software Engineering, ICSE’05, May 2005.
[5] IASRI, Association rule mining, http://www.iasri.res.in/ebook/win_school_aa/notes/association_rule_mining.pdf
[6] Ghosh, A., and Nath, B. 2004. Multi–objective rule mining using genetic algorithms. Information Sciences 163:123–133.
[7] Peter P. Wakabi-Waiswa Venansius Baryamureeba, Extraction Of Interesting Association Rules Using Genetic Algorithms International Journal of Computing and ICT Research, 2(1), 2008

