UML Models of Research Process in Empirical Software Engineering

Wasiur Rhmann

Dept. of CS and IT, Babasaheb Bhimrao Ambedkar University (A Central University) Satellite Campus, Amethi, India

*Corresponding Author: wasiurrhmann786@gmail.com,

Abstract— In recent years, Empirical software engineering is gaining popularity among the software engineering and research community to improve the software quality. However lack of clear understanding about how to do empirical research in software engineering poses threats of unreliable research results. |So to clear understand the process of doing research in empirical software engineering, Unified Modeling Language (UML) models are used. Unified Modeling Language in software engineering is used for designing and analysis of the software systems.

Keywords—Unified Modeling Langugae, Activity diagram, Use Case diagaram, State diagram, Empirical Software engineering

I. INTRODUCTION

With the increase in software size and complexity software organizations are facing problems of producing quality software. Earlier in the community of software professionals there was debate on is software is engineering discipline that debate in now over and Software engineering is well for systematic development of the quality software Traditionally software engineers use their experience and expertise to develop high quality software. They use software process and tools in the development of software without based on any scientific evidence. Empirical studies are verified with the help of observations. Empirical software engineering helps to plan, monitor and control and mange the ways in which software products are being developed. Different phases of the software development life cycle use different tools. Effectiveness of these tools can be evaluated by empirical studies. Now empirical software engineering comes to provide effectiveness about tools and process followed for the software development. Empirical term is related with the observations or experience [1]. Models of the system helps to better understand the system without creating it. Unified Modeling Language is a modelling language used to model the design of systems in software engineering. UML offers different diagrams to model the system and usually a set of diagram from UML is used to model the system. Software engineering uses Model driven development techniques to describe the system [2]. Modeling techniques which were initially developed to describe the behaviour of software system are now extensively used to describe the non-software process modelling. UML diagrams capture two different views of the system which are described as:

Static view: Static view of a system is expressed with the help of a diagram like class, object and these diagram present static structure of the system

Behavioural view: Interaction between objects are used to express the behavioural views of the system. Different diagrams like activity, state sequence, use case diagrams are used to express the behavioural view of the system. UML is also useful to model non-software systems like design of hardware, workflow in medical electronic, health care systems.

Since its inception UML has been evolved to be able to be use in different domains [3]. UML models are used in different fields like in performance comparison of relational and object oriented database [4], in test cases generation in software testing [5], in planning of town and country [6], performance evaluation of network system [7]. The reason to use process modelling was to improve understanding about research process in empirical software engineering.

Section II contains recent researches related to the empirical software engineering. Section III contains the background of Unified Modeling Language and Concepts related to the empirical software engineering. Section IV contains various models of the UML to explain the research process in the empirical software engineering. In this section we have elaborated how empirical study is performed in software engineering and finally section V concludes the work.

II. RELATED WORK

In this section recent work related to empirical software engineering and work related to the UML modelling are described. Bansal [8] performed an empirical analysis of search based algorithms to identify the change prone classes

authors constructed prediction models using machine learning and Search based algorithms. Authors constructed different research questions related to the change proneness and its relation with OO metrics. They statistically compared the machine learning and search based algorithms in the prediction of change proneness. Verner et al. [9] performed an empirical study on the investigation of motivating factors for software development teams. Authors have taken four country named as Australia, Chile, Vietnam and USA for identifying the motivating factors in the software development. Rhmann and Saxena [10] used UML sequence diagram to model the aadhaar card number based ATM system. Saxena and Arora [11] have used UML to compare the performances of Object oriented systems. Ahmad and Saxena[12] have used to model the Air traffic control system. They used different types of UML diagram to clearly explain about the functioning of the Air traffic control system. Ansari et al. [13] used UML state diagram to model the ATM system. Risk associated with different state is also modelled and test suite is prioritized with the risk severity. Ansari and Saxena[14] have modelled the traveller management system using Unified Modeling Language. They used different diagrams to describe the functioning of traveller management system is details.

III. BACKGROUND

A. Unified Modeling Language

UML is used to create the blueprints of the system. It offers a numbers of diagrams to create different views of the systems [15]. In our work we have used three types of diagrams namely Activity diagram, Use case diagram and State diagram. The characteristics of these diagrams are described below:

1. UML Activity diagram

An activity diagram represents dynamic and high level view of the system and is used to model different activities which are taking place during an operation of the system. It helps to understand business requirements clearly. It may contain serial as well as concurrent activities and also offers facility to perform decision at any activity node.

2. UML State diagram

UML state diagram represents the object's different states in which it goes during its lifetime and the events which trigger the transitions from one state to other state. States are represented with round ended rectangles and transitions from one state to other are represented by arrows.

3. UML Use Case Diagram

Use case diagram captures the system's view from users point. It depicts what the system will do and what the task can be performed by different users of the system. I uses actors and use cases do describe the functionality of the system. Actors of the system are persons which interact with the system to perform some task which use cases represents the tasks performed by the actors.

B. Components of Empirical Research

There are mainly four components of empirical research process namely: purpose, products, process and participants.
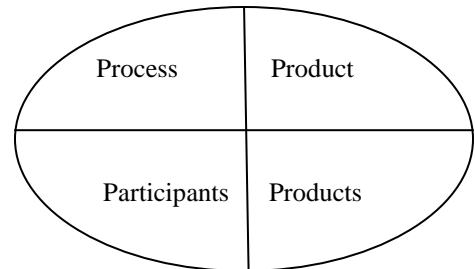

Figure 1 Component of Empirical Research

Purpose is the aim for which empirical research is being performed, products are the final outcome of the research and process is the techniques which are used to conduct research and participants are the people which are observed to obtain the research results.

C. Statistical Test

Statistical test helps to take decision about a process quantitatively. Statistical test verify whether there is enough evidence to reject an assumption or hypothesis. Hypotheses are used to define the relationship between the variables of the experiment. Two types are hypothesis are formulated to perform a statistical test:

Null hypothesis and alternate hypothesis
1. Null hypothesis: Null hypothesis is assumed to be true till it is supposed to be rejected by researchers
2. Alternate hypothesis: Alternate hypothesis is just opposite to the null hypothesis and alternate hypothesis becomes true when null hypothesis is rejected

Significance level (alpha) It refers to the probability of rejecting null hypothesis when it is true.

P-value is a probability value which determines whether the results obtained by statistical test are significant. Small p-value indicates that result is statistically significant.

D. Types of Statistical Test

Statistical tests which are applied for testing the validity of hypothesis are categorized in two categories:

Parametric test and Non parametric test
1. Parametric Test-These test are applied when data is normally distributed and it has bell shape. Example t-test, ANNOVA
2. Non-Parametric Test- These tests are applied when data is highly skewed. Example Kruskal–Wallis–Wilcoxon test

IV UML models to describe empirial research
In this section different UML models like UML activity diagram, Use case diagram and state diagram are used for explaining the process of doing empirical research in software engineering.

A. UML Activity diagram s for reserach process in empirical software engineering
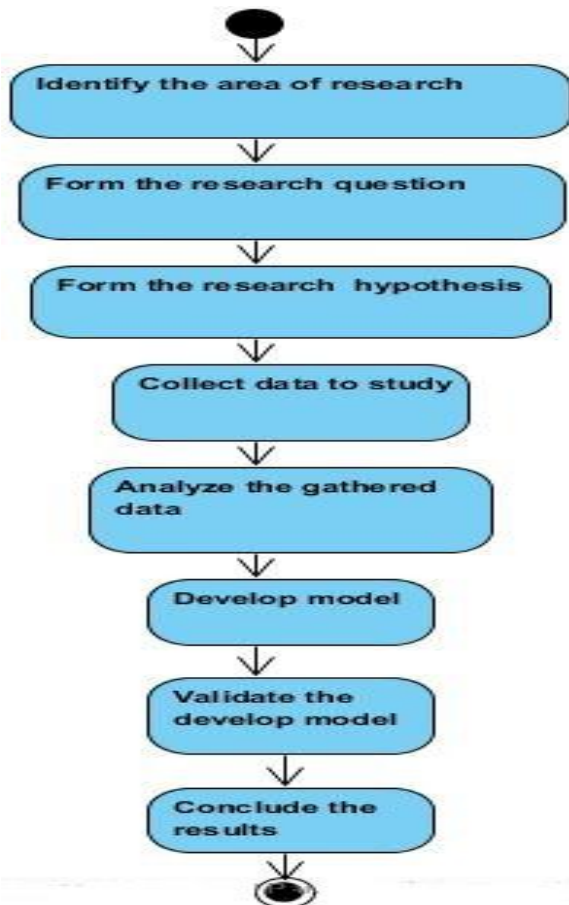

Figure 1 UML Activity Diagram for Empirical Study Process

Fig. 1 describes the steps which have to be followed to perform an empirical study in software engineering. First step to conduct an empirical research is to identify the area of research then form research questions which are supposed to be answered by the study. From the research questions formulate statistical hypothesis. Then data related to the study in collected either from primary or secondary sources. Collected data is analyzed statistically and a model is developed for prediction of dependent variable and developed model is validated with the data and finally results are reported.
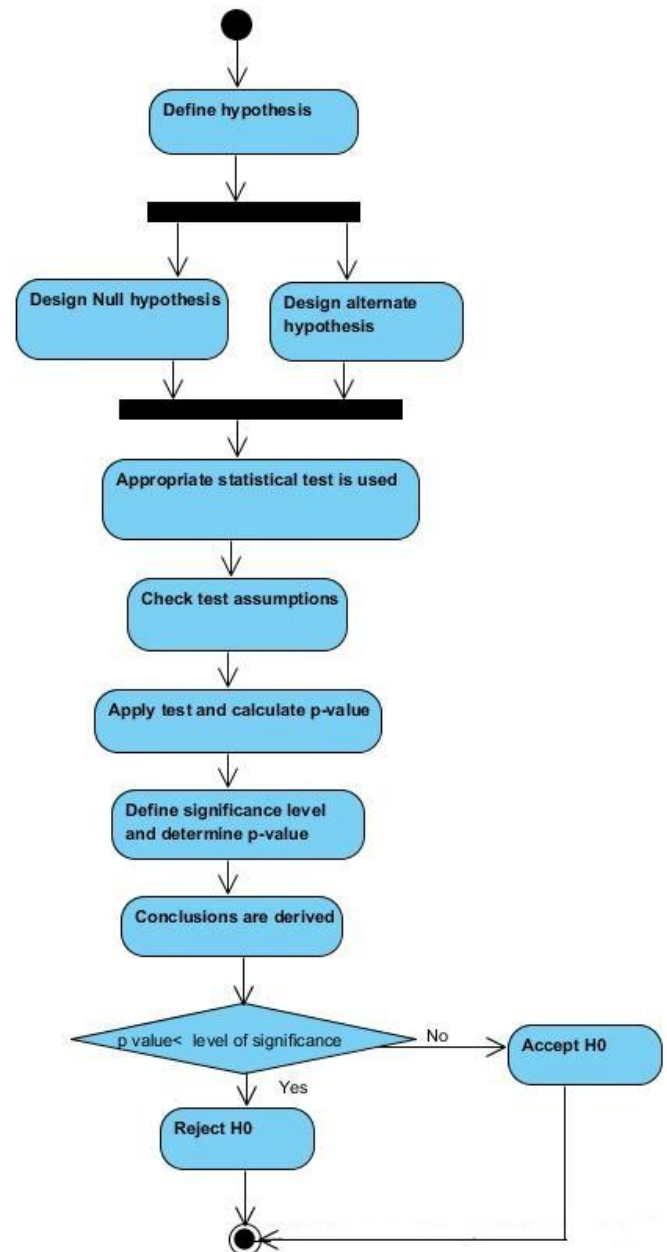

Figure 2 UML Activity Diagram for Hypothesis Testing

Fig. 2 shows UML activity diagram for Testing of Hypothesis. First we define the hypothesis and then create null and alternate hypothesis. After the formulation of research questions research hypothesis are created. Research hypothesis are formed on the basis of key variables of the study and the relationship between the key variables. State the null hypothesis and alternate hypothesis. Null hypothesis is expected to be rejected. For assessing the null hypothesis different statistical tests are used and applications of these statistical test tests are based types of distribution, sample size, number of variables in the study. P-value is calculated to determine the significance of the results. P-value is probability that if null hypothesis is true statistical summary

or results would be the same or greater than the actual observed results. Null hypothesis is rejected if the p-value is less than or equal to threshold value (alpha) which is determined by researcher and referred as level of significance.

B. UML Use Case Diagram for research process in empirical software engineering
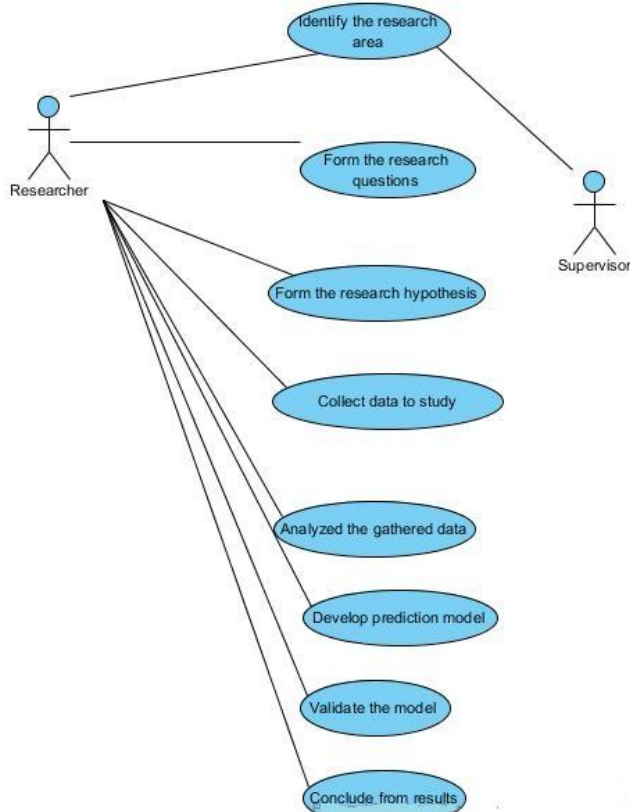

Figure 3. UML Use Case Diagram for Empirical Software Engineering Research

Fig. 3 describes the empirical software engineering research process in this diagram there are two actors' researcher and supervisor. Researchers do the activity of identification of research area with the help of supervisor. He forms the research questions, formulate the research hypothesis, collect data, analyze data, develop prediction models, validation model and conclude from results.

In the fig. 4 we have described the process of creation of a prediction model which predicts dependent variable. First state is gathering of data and then from gathered data independent variables are selected then we move to next state where reduced set of independent variables are selected as there may be some redundant variables in the independent dataset. These selected reduced set of independent variables are given to the model designed with the learning technique. Independent and dependent variables are used in next state to train the model after training model; prediction model is tested with the test data which gives performance measures

of the model to next state where we can compare performance.

C. UML State diagram for reserach process in empirical software engineeering
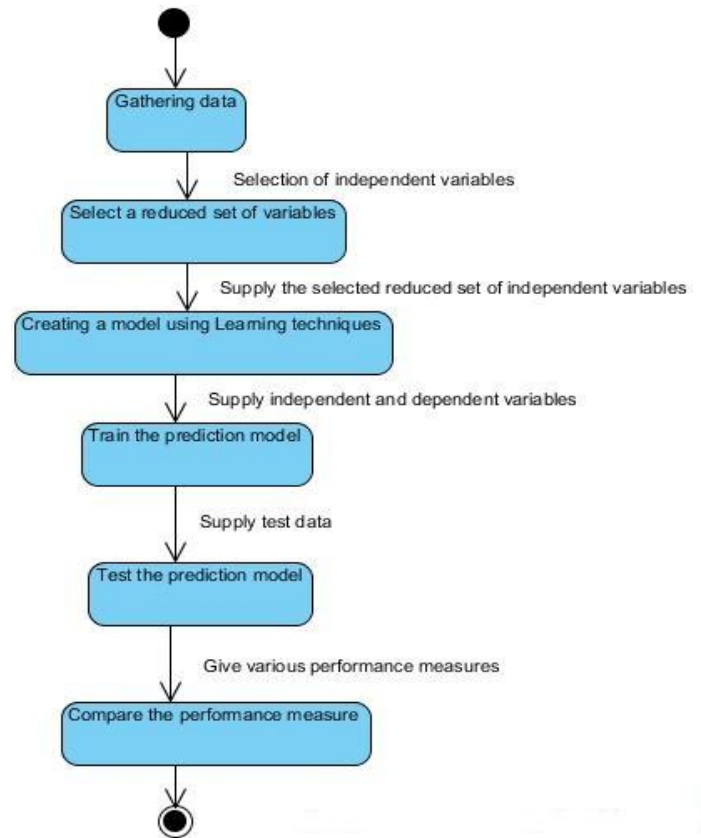

Figure 4. UML State Machine Diagram to Create a Model to Predict and Attribute

IV. CONCLUSION AND FUTURE SCOPE

UML models have been used by researchers to better understand the system before actually implementing and these models are used in various fields. In the present study, UML models are used to explain the empirical research process in software engineering. These UML models may be useful for novice researchers in the field of empirical software engineering. As the empirical software engineering is useful to assess the effectiveness of process, products and tools used in different phases of the software development life cycle, use of UML to understand the research process can improve the overall quality of the software.

REFERENCES

[1] R. Mahrotra, "Empirical Research in Software Engineering Concepts Analysis and Applications", CRC press Taylor and Francis Group, New York, 2016.

[2] I. Sommerville, "*Software Engineering*", 4th Ed.,MA. Addison Weslay, 1992.

[3] Aldowaison, TA, Cafar L. K. "*Business process reengineering: an approach for process mapping*". Omega International Journal of Management Science, Vol. 27, pp. 515-524, 1999.

[4] Vipin Saxena, Ajay Pratap, "*Performance comparison between relational and object-oriented databases*", International Journal of Computer Applications, 2013, Vol. 21, No. 22, pp. 6-9.

[5] W. Rhmann, V. Saxena, "*Optimized and Prioritized Test Paths Generation from UML Activity Diagram using Firefly Algorithm*", International Journal of Computer Applications, 2016, Vol. 145, No. 6, pp. 16-22.

[6] T. Zaidi, V. Saxena, "*A UML Frame Work for Town and Country Planning*", International Journal of Computer Applications, 2013, Vol. 73, No. 20, 2, pp. 23-25.

[7] Vipin Saxena, Deepak Arora, Manish Shrivastava, *"Performance evaluation of network system through UML*", ACM SIGSOFT Software Engineering Notes, 2009, Vol. 34, No. 5, pp. 1-3.

[8] A. Bansal,"*Empirical analysis of search based algorithms to identify change prone classes of open source software*", Computer Languages, Systems & Structures, 2017, Vol. 47, pp. 211–231.

[9] J. M. Verner, M. A. Babar, N. Cerpa, T. Hall, C. Beecham, *"Factors that motivate software engineering teams: A four country empirical study",* The Journal of Systems and Software, 2014, Vol. 92, pp. 115-127.

[10] W. Rhmann, V. Saxena, *"Test Case Generation from UML Sequence Diagram for Aadhaar Card Number based ATM System",* International Journal of Applied Information System, 2016, Vol. 11, No. 4, pp. 37-43.

[11] V. Saxena, D. Arora, "*Performance evaluation for object oriented software systems",* ACM Software Engineering Notes, 2009, Vol. 34, No. 2, pp. 1-5.

[12] S. Ahmad, V. Saxena, "*Design of formal air traffic control system through UML*", Ubiquitous computing and communication journal, 2008, Vol. 3, No. 6, pp. 11-20.

[13] G. Ansari, W. Rhmann, V. Saxena, "*Risk Based Test Case Prioritization using UML State Machine Diagram using Risk Information*", International Journal of Applied Information System, 2016, Vol. 11, No. 7, pp. 15-20.

[14] G. Ansari, V. Saxena, "*Object Oriented UML modeling for traveler Management System"*, Ubiquitous Computing and Communication Journal, 2008, Vol. 3, No. 3

[15] G. Booch, J. Ramaugh and I. Jacobson, *The Unified Modeling Language User Guide: Addison Wesley, 1999.*

.

Authors Profile

Mr. Wasiur Rhmann have completed his B. sc.(Hons) in Physics and M. C. A. for Aligarh Muslim University India. He has submitted his Ph. D. thesis in Department of Computer science, Babasaheb Bhimrao |Ambedkar University (A Central University, Lucknow India. He is working as Assistant Professor(Guest Faculty) in Department of Computer Science and Information Technology. His research interest include Software engineering, software testing and UML modeling. He have been published a number of papers in various International journals of repute.

.