A Survey on Using Artificial Intelligence Techniques in the Software Development Process

K. Hema Shankari¹, Dr. R. Thirumalaiselvi²
¹ Research Scholar, Bharath University and Assistant professor, Department of Computer Science, Women’s Christian College, Chennai.
² Research Supervisor and Assistant professor, Department of Computer Science, Govt. Arts College (Men) (Autonomous), Nandanam, Chennai.

ABSTRACT
Software engineering and artificial intelligence are the two important fields of the computer science. Artificial Intelligence is about making machines intelligent, while Software engineering is knowledge-intensive activity, requiring extensive knowledge of the application domain and of the target software itself. This study intends to review the techniques developed in artificial intelligence from the standpoint of their application in software engineering. The goal of this research paper is to give some guidelines to use the artificial intelligence techniques that can be applied in solving problems associated with software engineering processes. The aim of this paper is to find out the exact AI technique is likely to be fruitful for particular software development process

Keywords: Software Engineering, Artificial Intelligence Techniques, Software Development Process

I. INTRODUCTION
Software development process is a very complex process that, at present, is primarily a human activity. Programming in software development, requires the use of different types of knowledge: about the problem domain and the programming domain. It also requires many different steps in combining these types of knowledge into one final solution. There are various techniques in artificial intelligence (AI) from the standpoint of their application in software engineering that can be deployed in solving problems associated with software development processes. Artificial Intelligence is concerned with the study and creation of computer systems that exhibit some form of intelligence and attempts to apply such knowledge to the design of computer based systems that can understand a natural language or understanding of natural intelligence. Many Software products costs can be attributed to the ineffectiveness of current techniques for managing this knowledge, and Artificial Intelligence techniques can help alleviate this situation.

II. LITERATURE SURVEY: AI TECHNIQUES
Hany M Ammar, Walid Abdelmoez and Mohamed Salah (2012) in their paper on the Current state and open problems in the Software Engineering using artificial intelligence discuss about how the artificial intelligent technique such as KBS,CBR, Fuzzy logic and automated programming tool help to overcome the problems associated in the traditional software development. There are certain open problems such as SBST requires further research

Mark Harman (2011) discuss about the three boards areas of AI techniques such as SBSE, Fuzzy , probabilistic method , classification learning and prediction help the software engineering community and also about the challenges ahead in AI for SE.

Farid Meziane, Sunil Vadera (2010) discuss about the current developments and future prospects for Artificial Intelligence in software Engineering. Testing and the other phase of the software development.

Farah Naaz Raza (2009) in the paper “Artificial Intelligent technique in Software engineering (AITSE) “ explains about by using AI based systems with the help of automated tool or automated programming tool we can eliminate risk assessment phase by saving the time in software development and also AITSE reduce the development time in the software development.

Parveen Ranjan Srivastava and Tai-hoon Kim (2009) discuss about the application of Genetic Algorithm in Software Testing and this paper clearly says that the GA is used for the improvement of the software testing efficiently.

Mark Harman and Bryan F. Jones (2001) presented paper on the Search-based software engineering and in this paper explains about the search-based techniques could be useful for the development of software measures.
Prince Jain (2011) discusses about the interaction between Software Engineering and Artificial Intelligence and in this paper he explains about the reasons for which AI techniques are needed for the software engineering process.

Jonathan Onowakpo Godfrey Ebbah (2002) presented paper on Deploying Artificial Intelligence Techniques in Software Engineering in this paper intends to review the techniques developed in artificial intelligence (AI) from the standpoint of their application in software engineering. In particular, it focuses on techniques developed (or that are being developed) in artificial intelligence that can be deployed in solving problems associated with software engineering processes.

Jyoti tewari,Swati arya, Prem narayan singh (2013) presented paper on Approach of Intelligent Software Agents in Future Development discuss about the intelligent behavior of the intelligent agent predicting the future development in a software.

Dr. Nachamai,M. Senthil Vadivu and Tapaskar (2011) in the paper enacted software development process based on agent methodologies discuss about how agent based software development have weight factors more than traditional software models.

Seth Hock, (1989) in the paper Computers And Computing explains artificial intelligence on the other hand is a domain of computer science that attempts to make machines perform tasks that hitherto done by human beings.

SURVEY ON AI TECHNIQUES IN SOFTWARE CODING AND TESTING

Techniques learned from AI research make advanced programming much simpler, especially with regard to information flow and control as a result of advances in knowledge representation. In the following we focus on the AI techniques used in supporting the tasks of coding and testing.

a) Coding:

Software engineers can apply AI techniques to help automate or assist the programming process.

Use of AI to help automate the programming process:

The idea here is to have a completely automated program synthesis. This is done by having human specialists write a complete and concise specification of the desired software; so that, a system can generate "functions, data structures, or entire programs" directly from the specifications. There are many possible AI technologies that could be applied. Booch (1986) explains the NL description to data types while describing the Object-Oriented analysis and design method. Analogical reasoning in software reuse can be used. The idea is to find a system with similar requirements and modify it. Although this process looks feasible, it has not been demonstrated in software engineering to any great extent.

Closely related to analogical reasoning techniques is Case-based reasoning (CBR). CBR is based upon the premise that similar problems are best solved with similar solutions. CBR is argued to offer a number of advantages over many other knowledge management techniques. For program synthesis retrieval from component repositories and the reuse of successful past Experience is important. As an example, one application of CBR technology was to support the reuse of software packages within Ada and C program libraries.

The idea of experience reuse, the most ambitious form of CBR-supported reuse, is closely aligned with what is called Experience Factory. This field is also known as Organizational Learning, researches methods and techniques for the management, elicitation, and adaptation of reusable artifacts from software engineering projects. An Experience Factory is based upon a number of premises such as a feedback process, appropriate storage of experience, and support of reuse and retrieval.

Constraint programming is another AI technique that is applied in software engineering. Constraint programming has been, for example, used to design the PTIDEJ system (Pattern Trace Identification, Detection and Enhancement in Java. PTIDEJ is an automated system designed to identify micro-architectures looking like design patterns in object oriented source code.

A micro-architecture defines a subset of classes in an object-oriented program. The main interest of PTIDEJ is that it is able to provide explanations for its answers. This is really interesting since coding and software engineering is often considered a form of art and where fully automated systems are not always appreciated by potential users (or programmers).

Search Based Software Engineering (SBSE) is an emerging research topic that focuses on representing aspects of Software Engineering as problems that may be solved using meta-heuristic search algorithms developed in AI. SBSE is the reformulation of software engineering tasks as optimization problems. One of the optimization and search techniques that can be used are genetic algorithms. Genetic algorithms are used for automatic code generation by optimizing a population of trial solutions to a problem. The individuals in the population are computer programs.

b) Testing:

Software testing remains an expensive task in the development process and one of the main challenges concerns its possible automation. AI techniques can play a vital role in this regard. One of
these techniques are constraint solving techniques. Since the seminal work of Offut and De Millo (1991) in the context of mutation testing, much attention has been devoted to the use of constraint solving techniques in the automation of software testing (Constraint-based testing). ATGen, for example, is a software test data generator based on symbolic execution and constraint logic programming for ADA programs. There are many other ways how AI techniques can support the testing process. One of the earliest studies to suggest adoption of a knowledge based system for testing was by Bering and Crawford (1988) who describe a Prolog based expert system that takes a Cobol program as input, parses the input to identify relevant conditions and then aims to generate test data based on the conditions. A more active area of research since the mid-1990s has been the use of AI planning for testing. An AI planner could generate test cases, consisting of a sequence of commands by representing commands as operators, providing initial states, and setting the goal as testing for correct system behavior. AI planning was also used for testing distributed systems and for the generation of test cases for graphical user interfaces. Stuart Russell and Peter Norvig, (1994) in their paper Artificial Intelligence: A Modern Approach explain the Analysis is the process of breaking something into pieces or components with a view to understanding the individual components. A study by Kobbacy, (2007) has shown that the use of genetic algorithms for optimization has grown substantially since the 1980s. This trend is also present in their use in testing, with numerous studies aiming to take advantage of their properties in an attempt to generate optimal test cases. Bertolino (2007) presents a useful framework for summarizing the challenges that are faced in addressing the problems of ensuring that systems are fit for purpose, suggesting further research on: (i) developing a universal theory of testing, (ii) fully automatic testing, (iii) design to facilitate testing and (iv) development of integrated strategies that minimize the cost of repeated testing. Wappler and Wegener (2006) acknowledge that using a fitness function as the primary means of avoiding illegal sequences is not efficient. Instead they propose a novel use of Genetic Programming (GP), which aims to learn functions or programs by evolution. The authors in for example, used genetic algorithms for testing object oriented programs where the main aim was to construct test cases consisting of a sequence of method calls. Nand, S., Kaur, A., Jain S. (2007) in the paper Use Of Fuzzy Logic In Software Development. Issues in Information Systems. Explains about the use of fuzzy logic in software testing to manage the uncertainty involved in this phase of software development.

SURVEY ON AI TECHNIQUES IN PLANNING AND PROJECT EFFORT ESTIMATION

Good project planning involves many aspects: staff need to be assigned to tasks in a way that takes account of their experience and ability, the dependencies between tasks need to be determined, times of tasks need to be estimated in a way that meets the project completion date and the project plan will inevitably need revision as it progresses. AI has been proposed for most phases of planning software development projects, including assessing feasibility, estimation of cost and resource requirements, risk assessment and scheduling. This section provides pointers to some of the proposed uses of knowledge based systems, genetic algorithms, neural networks and case based reasoning, in project planning and summarizes their effectiveness.

Knowledge Based Systems

There have been several studies that adopt this assumption and aim to capture this experience in a Knowledge Based System (KBS) and attempt to utilize it for planning future software development projects. One of the earliest studies to suggest adoption of a Knowledge Based System (KBS) for testing was by Bering and Crawford (1988) who describes a Prolog based expert system that takes a Cobol program as input, parses the input to identify relevant conditions and then aims to generate test data based on the conditions. Sathi, Fox & Greenberg (1985) argue that a well defined representation scheme, with clear semantics for the concepts associated with project planning, such as activity, causation, and time, is essential if attempts to utilize KBS for project planning are to succeed. Hence, they develop a representation scheme and theory based on a frame based language, known as SRL (Wright, Fox, & Adam, 1984). Their theory includes a language for representing project goals, milestones, activities, states, and time, and has all the nice properties one expects, such as completeness, clarity and preciseness. Surprisingly, this neat frame based language and the semantic primitives they develop have been overlooked by others and appear not to have been adopted since their development. Gupta, Bastani,Khan & Yen (2004) take advantage of the goal oriented properties of Means-Ends planning by defining potential system actions as operators so that generating tests becomes equivalent to the goal of finding a plan from the current state to specified unsafe or near unsafe states.
Memon, Pollack & Soffa (1999) argue that human generation of test cases for graphical user interfaces requires enumeration of a large number of possible sequences of user actions, making the process inefficient and likely to be incomplete. Instead, as with the above studies, they propose the use of production rules and associative networks which seemed promising at the time have not been widely adopted. Boardman, J. T., & Marshall, G. (1990) explains about the knowledge-based architecture for project planning and control. When considering whether to adopt a KBS approach, the cost of representing the knowledge seems high and unless this can be done at a level of abstraction that allows reuse, one can imagine that it is unattractive to software developers who are keen and under pressure to commence their projects without delay.

**Neural Networks**

Neural networks (NNs) have been widely and successfully used for problems that require classification given some predictive input features. They therefore seem ideal for situations in software engineering where one needs to predict outcomes, such as the risks associated with modules in software maintenance and software risk analysis (Neumann, 2002) and for predicting faults using object oriented metrics (Thwin & Quah, 2002). Karlsson and Ryan (1997), Khoshgoftaar & Lanning (1995) explains the neural network approach. The study by Hu, Chen, Rong, Mei & Xie (2006) is typical of this line of research. They first identified the key features in risk assessment based on past classifications such as those presented by Wallace and Keil (2004) and further interviews with project managers. They identified a total of 39 risk factors which they grouped into 5 risk categories: project complexity, cooperation, team work, project management, and software engineering. These were reduced to 19 linearly independent factors using principal component analysis (PCA). Projects were considered to have succeeded, partially failed, or failed. In their experiments, they tried both the use of a back propagation algorithm for training and use of GAs to learn networks, using 35 examples for training and 15 examples for testing. The accuracy they obtained using back propagation was 80% and that with a GA trained NN was over 86%, confirming that use of NNs for predicting risk is a worthy approach, though larger scale studies are needed.

**Genetic Algorithms**

There have been numerous uses of genetic algorithms for project scheduling in various domains (Cheng & Gen, 1994; Hindi, Hongbo, & Fleszar, 2002; Hooshyar, Tahmiani, & Shenasa, 2008; Yujia & Chang, 2006; Zhen-Yu, Wei-Yang, & Qian-Lei, 2008, Briand, Labiche, & Shousha, 2005). A survey of their application in manufacturing and operations management can be found in (Kobbacy, Vadera, & Rasmy, 2007; Meziane, Vadera, Kobbacy, & Proudlove, 2000). These typically formulate project planning as a constraint satisfaction problem with an objective that needs optimisation and, which is then transformed into a form suitable for optimisation with a GA. A study by Kobbacy, Vadera and Rasmy (2007) has shown that the use of Genetic Algorithms (GAs) for optimization has grown substantially since the 1980s and this growth has continued while the use of other AI technologies has declined. This trend is also present in their use in testing, with numerous studies, trying to take advantage of their properties in an attempt to generate optimal test cases (Baresel, Binkley, Harman, & Korel, 2004; Baudry, Fleurey, Jezequel, & Le Traon, 2002a, 2002b; Briand, Feng, & Labiche, 2002; Briand, Labiche, & Shousha, 2005; T. Menzies (2001) & Emrich & Lylod (1988) Wappler & Wegener, 2006). In the area of software development, Shan, McKay, Lokan & Essam (2002) utilize Genetic Programming to evolve functions for estimating software effort. Two target grammars were adopted for the functions that allowed use of a range of mathematical functions (e.g., exp, log, sqrt) as well as a conditional expressions. The approach was tested on data consisting of 423 software development projects characterized by 32 attributes (e.g. such as intended market, requirements, level of user involvement, application type, etc) from the International Software Benchmarking Standards Group (www.isbsg.org.au) with roughly 50% used for training and 50% used for testing. The results of this study show that the approach performs better than linear and log regression models. An interesting finding of the study was that although the most accurate functions discovered by GP utilized similar parameters to the traditional estimates, a key difference was that it adopted non-linear terms involving team size. Creating a good assignment of staff to tasks and producing schedules is critical to the success of any software development project. Yujia & Chang (2006) show how it is possible to utilize GAs to produce optimal schedules and task assignments. Their proposal involves a two part chromosome representation. One part includes the assignment of individuals to tasks and another involves representing the topological ordering of the tasks in a way that ensures that the offspring
generated using the cross-over operator remain valid schedules. The fitness function is obtained by utilizing a systems dynamics simulation to estimate expected task duration given a particular chromosome. The results of their experiments suggest that this is a promising approach, though further work on how to utilize GAs in practice when schedules change is still needed. An important part of developing an optimal schedule that meets a target completion date is the trade-offs that may occur. For example, attempts at increasing quality can result in increasing cost and possibly compromising completion time but perhaps increasing user satisfaction. Increasing resources on tasks increases the local cost but may result in early completion, higher quality and reduction of overall cost. Hooshyar, Tahmani & Shenasa (2008) propose the use of GAs to optimize schedules to take account of such trade-offs. They represent a schedule by a chromosome consisting of the activity duration and which is ordered based on their dependency. In their experiments, they utilize the standard mutation and two-point cross-over operators and adopt a fitness function that includes the cost and duration. The experimentation is carried out on projects consisting of 10, 20 and 30 activities and concludes that although the well known algorithm due to works well for small scale problems, GAs may be more effective for larger scale problems. Ryan (2000) explains about the Automatic re-engineering of software using genetic programming and Siemens (1971) explains the CPM time-cost tradeoff algorithm which are very useful.

Case Based Reasoning
It can be argued that successful project planning and management is heavily based on experience with past cases. It is therefore surprising that there are few studies that propose the use Case Based Reasoning (CBR) for project planning of software development. One of the few exceptions is the study by Yang and Wang (2009), who explore the combined use of CBR and data mining methods for project planning. They use a structured representation for cases, called Hierarchical Criteria Architecture (HCA), where projects are described in terms of the customer requirements, project resources and keywords describing the domain. The use of HCA enables different weights to be adopted when matching cases, allowing greater flexibility depending on the preferences of the project manager. Given a new project, first similar new cases are retrieved. Then, data mining methods, such as association rule mining, are used to provide further guidance in the form of popular patterns that could aid in project planning. In a trial, based on 43 projects, Yang & Wang (2009), show how the combined use of CBR and data mining can generate useful information, such as “the duration of project implementation was about 26 days and 85% of projects of projects were completed on time”, which can be used to provide guidance when planning a similar project.


III. TECHNIQUES AND TOOLS OF AUTOMATED PROGRAMMING
Because of the evolutionary nature of software products, by the time coding is completed, requirements would have changed (because of the long processes and stages of development required in software engineering): a situation that results in delay between requirement specification and product delivery. There is therefore a need for design by experimentation, the feasibility of which lies in automated programming. Some of the techniques and tools that have been successfully demonstrated in automated programming environments include:

• Language Feature: this technique adopts the concept of late binding (i.e. making data structures very flexible). In late binding, data structures are not finalized into particular implementation structures. Thus, quick prototypes are created which result in efficient codes that can be easily changed. Another important language feature is the packaging of data and procedures together in an object, thus giving rise to object-oriented programming: a notion that has been found useful in environments where codes, data structures and concepts are constantly changing. Lisp provides these facilities.
• Meta Programming: this concept is developed in natural language processing (a sub field of AI). It uses automated parser generators and interpreters to generate executable lisp codes. Its use lies in the modeling of transition sequences, user interfaces and data transformations.
• Program Browsers: these look at different portions of a code that are still being developed or analyzed, possibly to make changes, thus obviating the need for an ordinary text editor. The browser understands the structures and declarations of the program and can focus on the portion of the program that is of interest.
• Automated Data Structuring: this means going from a high-level specification of data structures to a particular implementation structure.

When systematic changes need to be made throughout a code, it is more efficient and controllable to do it through another program (i.e.,
program update manager) than through a manual txt editor. For instance, a change in program X may be required whenever h is being updated by b-1 under the condition that b is less than C. Assume that a program W makes a systematic change in all such places. If another program makes a change in W, then any program changed by W also must be updated. Thus, program update managers propagate changes. Because of this ability, program update managers are useful when prototypes need to be developed quickly.

IV. NEED FOR ARTIFICIAL INTELLIGENCE TECHNIQUES IN SOFTWARE ENGINEERING

Based on the above literature survey the most common reasons for which AI methods, tools and techniques are applicable to SE are discussed:

- Automatic Programming (AP) in AI is synonymous with Software Engineering and this represents a new paradigm for SE in the future research.
- Expert systems technology is sufficiently successful and mature enough to provide significant solutions to certain aspects of the SE process and problem.
- AI development and maintenance environments are suitable for direct application to the SE process.
- AI methodology and techniques can be applied to the software design process.
- The AI rapid prototyping model is useful as a SE paradigm.
- AI techniques reduces cost.
- Errors detected in coding will be isolated in the requirements stage.
- Changes need be made only at the requirements stage.

V. VARIOUS ARTIFICIAL INTELLIGENCE TECHNIQUES

The various Artificial Intelligence Technique used in the Software Development Process are listed below:

<table>
<thead>
<tr>
<th>AI Technique</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge Based System</td>
<td>Used in the design phase of the software development process</td>
</tr>
<tr>
<td></td>
<td>It manages the requirement phase , planning and project effort estimation</td>
</tr>
<tr>
<td>Neural Network</td>
<td>Eliminates the risk associated with modules in software maintenance and Used in the software engineering prediction outcomes.</td>
</tr>
<tr>
<td>Fuzzy logic</td>
<td>Reasoning the uncertainty</td>
</tr>
<tr>
<td>Genetic algorithm</td>
<td>Used in the software testing and generating test cases</td>
</tr>
<tr>
<td>Case Based Reasoning(CBR)</td>
<td>Used for finding out the duration or time taken to complete a project</td>
</tr>
<tr>
<td>Natural Language Processing</td>
<td>It helps in the user requirements and improves the phase of software development life cycle.</td>
</tr>
<tr>
<td>SBSE</td>
<td>Reformulating the software engineering problems as optimization problems</td>
</tr>
<tr>
<td>Rule induction</td>
<td>Used to defect prediction</td>
</tr>
</tbody>
</table>
VI. CONCLUSION AND SUGGESTIONS:
In this paper the promising research work on applying AI techniques to solve some of the most important problems facing the software engineers is studied. The study suggests that there is now good progress in the use of AI techniques in SE. Furthermore, the development of new areas such as intelligent agents and their use in distributed computing, context aware and secure applications will require closer links between SE and AI in the future.

The important suggestions emerge from the above survey findings are Artificial Intelligence techniques applied to the software engineering process can have a major impact on reducing the time to market, cost of development and enhance the quality of software system and used to support the tasks of coding and testing. Artificial Intelligence techniques are well suited to the complex software engineering problems, because they are designed to deal with the most demanding challenges. Artificial Intelligence based systems with the help of automated tool or automated programming tool time can be saved in the software development process.

REFERENCES

Table.1 Various Artificial Intelligence Techniques and its purposes

| Table.2 Artificial Intelligence Learning methods |
|-----------------|-----------------|
| S.NO | AI LEARNING METHODS |
| 1 | Failure Driven Learning |
| 2 | Learning by being Told |
| 3 | Learning by Exploration |

Table.3 Approaches of Artificial Intelligence

<table>
<thead>
<tr>
<th>S.NO</th>
<th>APPROACHES</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>classical approach</td>
<td>designing the AI</td>
</tr>
<tr>
<td>2</td>
<td>connectionist approach</td>
<td>letting AI develop</td>
</tr>
</tbody>
</table>

It uses the knowledge to overcome the risk management strategies during the software development process.

It develop automatically generate computer program and save the time in the coding phase.

Use for system redesign. It changes the traditional software development to expert system development.

Generation of program by computer usually based on specification.

Dealing with uncertainty

It generate new intelligent software system for better communication

Used in the field of engineering

Dealing with uncertainty
methodologies",IJEST , VOL.3, NO.11, November 2011.


