International Journal of Pure and Applied Mathematics
Volume 119 No. 14 2018, 1335-1339
ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue

ijpam.eu

A STUDY ON AGILE SOFTWARE DEVELOPMENT WITH PREDICTIVE DATA MINING & BIG DATA ANALYTICS

T.Sasi Vardhan[1], V.Anitha[2], Dr.C.V.P.R.Prasad[3]
[1]Research Scholar, Dept of CSE, UOT, Jaipur, India.
[2]Assistant Prof, BVRIT, Narsapur,India.
[3]Prof., & HOD, Dept of CSE MRECW, Hyderabad, India.

ABSTRACT—This paper proposes an agile model-based systems engineering (SE) methodology to engineer the contemporary large, complex, and interdisciplinary systems of systems. This paper introduces the reader the background of Big Data Analytics and how efficiently Agile methodology can be applied to achieve the business goal. As the world becomes increasingly dynamic, the traditional static modeling may not be able to deal with it. One solution is to use agile modeling that is characterized with flexibility and adaptability. On the other hand, data mining applications require greater diversity of technology, business skills, and knowledge than the typical applications, which means it may benefit a lot from features of agile software development. In this paper, we will propose a framework named ASD-DM based on Adaptive Software Development (ASD) that can easily adapt with predictive data mining applications. A case study in automotive manufacturing domain was explained and experimented to evaluate ASD-DM methodology.

Index Terms—Big Data Analytics, Agile, Big Data, Model Based System Engineering(MBSE), Agile Methodology, Adaptive Software Development.

INTRODUCTION

The Modern world is characterized by large interdisciplinary complex socio technical systems made of other system, personnel, hardware, software, information, processes, and facilities.The model-based SE (MBSE) paradigm is an emerging approach in the SE field. This approach relies on the development of a unified coherent system model that should act like a shared working platform and should reflect the stakeholders' ideas and positions in order to lead to a resulting successful system.

Data mining is the search for relationships and distinct patterns that exist in datasets but are "hidden" among the vast amount of data. A data mining task involves determining useful patterns from collected data or determining a model that fits best on the collected data. Although the idea of applying data mining techniques on software engineering data has existed since mid 1990s, only lately the idea has especially attracted a large amount of interest within software engineering . Data mining techniques are applied to analyze the problems raised during the life cycle of a software project development also to determine if two software components are related or not They were also used for software maintenance software testing , software reliability analysis and software quality.

Many questions arise when trying to apply data mining techniques on software engineering field. What types of SE data are available to be mined?, which SE tasks can be held using data mining?, how are data mining techniques used in SE? are all important questions that a lot of researches were trying to find, have relevant responses.

In this paper, we will focus on using agile modeling for predictive data mining applications, focusing on ASD (Adaptive Software Development) modeling, which replaced the static Plan-Design-Build lifecycle, with Speculate-Collaborate-Learn lifecycle. The main characteristics of ASD lifecycle are the continuous learning, intense collaboration among developers, testers, and customers, and it can easily adapt with uncertain future [17].

We will start by viewing the characteristics of data mining applications, and the most widely methodology used for process modeling for data mining applications (CRISP-DM methodology), then we will present the characteristics of agile modeling, and suggest a new framework named ASD-DM for data mining processes using Adaptive Software Development (ASD) method. The new framework was tested using a case study in the automobile manufacturing domain.

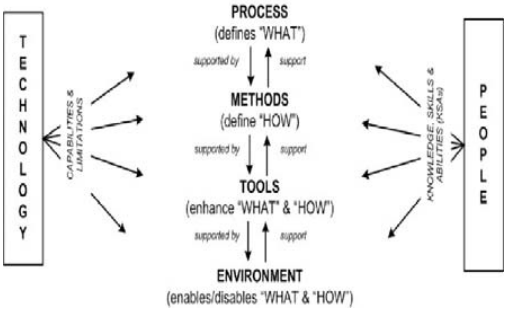
1. BACKGROUND STUDY: BIG DATA & DATA MINING APPLICATION

The amount of digital data is set to grow to 44 times its current volume by 2020. Big data arrives through different velocity, volume and variety. 2.7 Zeta bytes of data exist in the digital universe today. IBM estimates 2.5 quintillion bytes of data are generated each day. Ninety percent of the data in the world is less than two years old. Facebook stores, accesses, and analyses 30+ Pet bytes of user generated data. More than 5 billion people are calling, texting, tweeting and browsing on mobile phones worldwide. YouTube users upload 48 hours of new video every minute of the dayData mining applications are characterized by the ability to deal with the explosion of business data and accelerated market changes. These characteristics help providing powerful tools for decision makers. Such tools can be used by business users (not only PhDs, or statisticians) for analyzing huge amount of data for patterns and trends [19].

The most widely used methodology when applying data mining processes is named CRISP-DM[1]. It was one of the first attempts towards standardizing data mining process modeling [18]. CRISP-DM has six main phases, starting by business understanding that can help in converting the knowledge about the project objectives and requirements into a data mining problem definition, followed by data understanding by performing different activities such as initial data collection, identifying data quality problems, and other preliminary activities that can help users be familiar with the data.

The next and important step is data preparation by performing different activities to convert the initial raw data into data that can be fed into modeling phase. This phase includes tasks such as data cleansing and data transformation. Modeling is the core phase which can use a number of algorithmic techniques (decision trees, rule learning, neural networks, linear/logistic regression, association learning, instance-based/nearest-neighbor learning, unsupervised learning, and probabilistic learning, etc.) available for each data mining approach, with features that must be weighed against data characteristics and additional business requirements.The final two modules focus on evaluation of module results, and deployment of the models into production. Hence, users must decide on what and how they wish to disseminate/deploy results, and how they integrate data mining into their overall business strategy [18, 19].

A methodology can be seen as "a set of related activities, techniques, and conventions that implement one or more processes and is generally supported by a set of tools" .The process is the set of interacting activities that transform the inputs into outputs (the WHAT activities to perform), the method specifies the techniques for performing the tasks of the process (the HOW to execute), and the tools are the resources applied to the method in order to improve the efficiency of the tasks (enhancing the WHAT and the HOW) An MBSE methodology gathers all these pieces, implementing a given process, which is supported by a given method, which is enhanced by a set of tools.



2. USAGE OF BIG DATA

To determine more accurate information validated from different reliable sources

- To improve business decisions based on the real time data to increase revenue
- To improve marketing strategy and targeting right people
- To increase customer base
- To adapt to latest technology or trend.
- To increase the security and to avoid data leakage.

3. ASD METHOD ON PREDICTIVE DATA MINING APPLICATIONS: ASD-DM METHODOLOGY

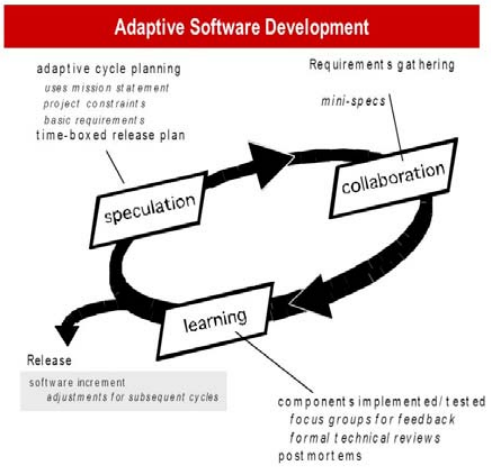
Software is intangible and more easily adapted than a physical product. Also, software processes depend on how a firm competes, and may be more adaptable than manufacturing processes bound by machinery, raw materials, and physical plants.
Technologies such as agile methods may make it less costly to customize and adapt development processes. Agile modeling has many process centric software management methods, such as: Adaptive Software Development (ASD), Extreme Programming (XP), Lean Development, SCRUM, and Crystal Light methods. Adaptive approaches are best fit when requirements are uncertain or volatile; this can happen due to business dynamism, and rapid evolving markets. It's difficult to practice traditional methodologies in such unstable evolving markets [11]. ASD modeling is one of such adaptive approaches. It replaces the static Plan-Design-Build lifecycle, with the dynamic Speculate-Collaborate-Learn life cycle. Speculation recognizes the uncertain nature of complex problems such as predictive data mining, and encourages exploration and experimentation. Predictive data mining problems require a huge volume of information to be collected, analyzed, and applied; they also require advanced knowledge, and greater business skills than typical problems, which need "Collaboration" among different stakeholders, in order to improve their decision making ability. That decision making ability depends on "Learning" component in order to test knowledge raised by practices iteratively after each cycle, rather than waiting till the end of the project. Learning organizations can adapt more easily with ASD life cycle.Hence the core of ASD is the premises were outcomes are naturally unpredictable, therefore, planning is a paradox. It is very difficult to successfully plan in a fast moving and unpredictable business environment, which is one of the main characteristics of predictive data mining application.

This is one of the major points that we are using to create a data mining process framework based on ASD methodology (figure1). We call our new methodology ASD-DM as it combines the characteristics of ASD methodology, with the prediction data mining solution steps.

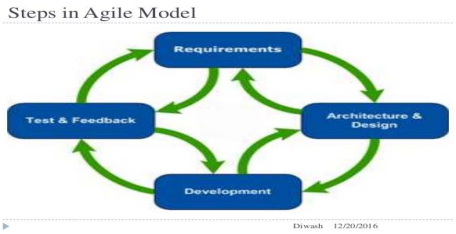
Speculation phase includes business and data understanding, and data preparations including ETL (Extract/Transform/Load) operations. This phase is the most important one as it takes considerable time and resources. This preparation phase will end by creating the enterprise data warehouse, and the required data marts and cubes.

Collaboration phase ensures the high communication in a diversity of experienced stakeholders in order o use the best modeling algorithm for predicative data mining process. Testing and evaluation of such algorithms occur in the "Learning" phase, the results will be discussed among the members of the project team, if the results are acceptable, a new release can be deployed in a form of predictive scoring reports, otherwise a new collaboration phase will be used in order to chose better data mining algorithm.

The cyclic nature of the whole framework can respond to the business dynamic changes, a new data sources can be added to the preparation phase, and the cycle will move again



Agile modelling: Agile Modelling (AM) is a practice-based methodology for modelling and documentation of software based systems. It is intended to be a collection of values, principles, and practices for modelling software that can be applied on a software development project in a more flexible manner than traditional modelling methods. The aim is to keep the amount of models and documentation as low as possible.



4. BIG DATA SOURCE

- Internal Company Database or Data warehouse
- Internal documents such as word document, excel sheet, csv files etc
- External data that flows into the company such as invoices, proposals, information from third party vendors etc.,
- Social websites and data feeds from twitter, facebook etc.,
- Message forums, public website, Google ads etc.,
- Customer information from different public websites
- Data from mobile apps

5. IN CONCLUSION

The development of an integrated methodology with simple, lean, and customizable processes and methods is of paramount importance to enable the widened utilization of MBSE practices. The SE process (WHAT) must be intuitive, logical, universal, and easy to use and to tailor. The ISO/IEC 15288 process standard requires some integration that can be provided by the SIMILAR process model. The MBSE method specifies the HOW to execute the process and relies on the development of a coherent system model. In this paper, we explained the use of data mining techniques in software engineering tasks such as programming, testing, maintenance, reliability, and quality. Due to the uncertain nature of predictive data mining application requirements, we proposed a new framework ASD-DM based on agile methodology, specifically Adaptive Software Development (ASD) methodology .This framework ensures continuous learning, and intense collaboration among developers, testers, and data mining customers.

REFERENCES

[1] A.E. Hassan, A. Mockus, R.C. Holt, and P.M. Johnson, “Guest editor’s introduction: Special issue on mining software repositories”. IEEE Trans. Softw. Eng., 31(6):426– 428, 2005

[2] J.C. Riquelme, M. Polo, J.S. Aguilar-Ruiz, M. Piattini, J. Francisco and F.T. Francisco-Ruiz, “A Comparison of Effort Estimation Methods for 4GL Programs: Experiences with Statistics and Data Mining”, International Journal of Software Engineering and Knowledge Engineering, Vol. 16, No. 1 (2006) 127-140.

[3] R. Nayak, T. Qiu, “A Data Mining Application: Analysis of Problems Occurring During A Software Project Development Process”, International Journal of Software Engineering and Knowledge Engineering, Vol. 15, No. 4 (2005) 647-663.

[4] Russell Jurney: Agile Data Science (2013)

[5] Mary Poppendieck and Tom Poppendieck. Lean software development: an agile toolkit. Addison-Wesley, Boston, 2003.

[6] Viktor Mayer-Schonberger, Kenneth Niel Cukier: Big Data: A Revolution That Will Transform How We Live, Work, and Think (2013).

[7] Mary Poppendieck and Tom Poppendieck. Leading lean software development: results are not the point. Addison-Wesley, Upper Saddle River, NJ (2010).

[8] M. Last, M. Friedman, and A. Kandel, “Using Data Mining for Automated Software Testing”, International Journal of Software Engineering and Knowledge Engineering, Vol. 14, No. 4 (2004) 369-393.

[9] M.K. Mattsson, N. Chapin, “Data Mining for Validation in Software Engineering: an Example”, International Journal of Software Engineering and Knowledge Engineering Vol. 14, No. 4 (2004) 407-427.