Software Product Quality Management Methodology & the Quantitative Assessment of Analyzability Indicators

Boumedyen Shannaq, Richmond Adebiaye

Abstract — The research paper developed a new software metric methodology for evaluating the analyzability indicator for software products. The proposed research methodology provided an objective and quantitative assessment in accordance with the requirements, limitations, purpose and specific features of software products. Forty-one (41) java programs were analyzed to extract and evaluate the software metrics described in Halstead metrics. The mathematical classification model was developed to replace the expert output in the evaluating process as related to the software metric indicators. The output of the algorithm was applied to identify the metrics with the greatest analyzability influence. The result indicated that 13 measured metrics with 98% of "analyzability" are relevant to seven (7) software code metrics with the remaining six (6) metrics making up only ~ 5% of "analyzability". The analyzed ROC-curves were similarly computed to test the performance of the proposed methodology compared to the expert's metric evaluation. The ROC-curves indicator for the proposed methodology showed resultant scores of ROC = 7.4 as compared to 7.3 from the experts' evaluation. However, both methods were correlated effectively after analytical computations with a resultant performance which showed that the proposed method outperforms the expert's evaluation.

Keywords — Software Metrics; Software Quality; Quantitative Assessment; Analyzability

I. INTRODUCTION

The development of programming has led to the creation of an entire industry for the production of software products. A software product today is an object of industrial production from a community of people that are involved in its creation, as the circulation of such products reaches hundreds of millions of copies [1]. The range of tasks that use software products is expanding. It is also obvious that a favorable situation has been created for the deeper integration of the software product in all spheres of human life [2]. On the way to the rapid development of programming, the transition from unit production to industrial has created a number of unresolved problems, one of which is quality management of the software product. The impossibility of achieving reliable and considerable quality control is due to the fact, that at present there are no formalized, quantitative methods for measuring the current level of quality, or ways to manage the software process. Obviously, this state of concern is undesirable in the industrial production of a product, since it is an additional risk factor that includes failure to meet production deadlines, exceeding the budget, etc [3][4]. In addition, the level of confidence in software products is growing unreasonably quickly, creating huge reliance and trust in areas of human activity associated with high responsibility including human survival [5]. Currently, the most common and used international standard in the field of software product quality is ISO / IEC 9126 [6]. Fig.1 describes the 'Maintainability' metrics with the new software sub-characteristics.

Fig.1 Maintainability metrics [6]

The quality model defined in this standard is taken as the basis for this research. The presence of inconsistency determines the existence of an urgent scientific and technical problem, leading to the urgent need to develop methods for quantitative assessment of software quality. This research aims to develop a methodology for quality indicators preferably - "analyzability" using quantitative assessment.

II. BACKGROUND AND LITERATURE REVIEW

Many large international standardization organizations deal with the problem of evaluating and ensuring the quality of software products, which emphasizes the existence and relevance of the problem. To date, several dozens of international standards have been issued that regulate issues related to the quality of software products [1]. Currently, the most common and used international standard in the field of software product quality is ISO / IEC 9126 which was reviewed in [6]. The quality model defined in this standard is taken as the basis for this study. Today, [2] the "analyzability" of the program code refers to quality indicators of software quality and is measured using the expert’s judgment.
However, such an approach to measuring quality indicators has a number of significant drawbacks: subjectivity of the assessment and conscious subjectivity is possible, i.e. interest in making any decision, the high cost of such an assessment, the absolute repeatability of the results and finally, the inability to conduct an assessment due to the lack of members of the expert council. Obviously, there are needs for new methods for quantitative assessment as the basis of quality indicators for the measurement. [6] In the field of software development, software quality assessment is still in its transitional stage and has yet to be fully implemented. At this stage in the development of programming, there is no acceptable description of the quality of software. There is also no common understanding of what quality should consist of or other ways to implement quality measurement. [7][8]

There are many definitions of the concept of quality, which, in essence, boil down to the totality of the technical, technological and operational characteristics of the products or processes, by which they are able to meet and satisfy the requirements of the customer. Currently, the most common and used international standard in the field of software quality assessment is ISO / IEC 9126 which defines quality as “the entire scope of signs and characteristics of a product or service that relates to their ability to satisfy the established or its anticipated needs”. In the ISO / IEC 9126 standard, six basic quality characteristics are given, of which, today, only a small part can be measured quantitatively. While the remaining quality indicators relate only by measurement accuracy, to qualitative or categorical-descriptive analysis. Obviously, this state of affairs leads to unreasonable management decisions in the field of quality management made on the basis of informal, intuitive quality assessments, which leads to unjustified risk, failure to meet project deadlines, exceeding budgets and, ultimately, to financial losses [9]. [10] expanded in related research that machine learning algorithms could be used to predict software bugs. [11] reviews many kinds of literature and formulate a description of the software quality. [12] proposed new use of information flow tracing to confuse the control flow of a program, to overthrow hateful code injection when considering the code size and the time. [13] proposed model LCC and LCOM measurements as cohesion metrics. Chidamber and Kemerer coupling metrics were used for assessing coupling. Their experiments found that measuring class cohesion based on totally shared attribute practice and method call is not sufficient. [15] proposed an empirically model for quantifying analyzability using object-oriented software. In their study, they found that the Strategy complexity of software is also an influencing factor of analyzability but with a negative impact. [16] discusses numerous metrics in each of five types of software quality metrics: product quality, in-process quality, testing quality, maintenance quality, and customer satisfaction quality.

III. ALGORITHM FOR ANALYZABILITY

The presentation of the algorithm could be different in terms of verbal description, representation appearance, description in pseudo-code, and etc. Each algorithm, from the point of view of human analysis, has a complexity that can be expressed in time that a specialist needs to spend in order to understand the principles of the algorithm. If any algorithm inherently already has initial complexity, it is obvious that different algorithms will have different initial complexity. It is also likely that the algorithm for adding two numbers will have less initial complexity than the algorithm for solving the traveling salesman problem. However, in the process of implementing the algorithm, some more factors are added to its initial complexity. The algorithm refers:

Let IC be the initial complexity of the algorithm.
Let ‘cof’ be a coefficient depending on the presentation form an algorithm of a description language.

Let IC’ be the complexity of the same algorithm for a specific implementation,

Therefore, the Expression will take the form:

\[
(cof \times IC’) / IC > 1
\]  (1)

It can be seen from Expression (1) that if the presentation form of the algorithm does not complicate its understanding (cof = 1), and the implementation does not complicate it, then the complexity of a particular implementation matches with the initial complexity of the algorithm. It is necessary to introduce a correction for the complexity of perception of the presentation form itself, i.e. the implementation of an algorithm using different presentation forms (for example, in different programming languages) will have different complexity, cof1 x IC’, cof2 x IC’ etc. The same description of the algorithm using the same presentation form can also have different complexity, i.e. time spent by a specialist on his understanding will be different cof1 x IC1’, cof2 x IC2’. This difference in the complexity of specific implementation depends on the quantitative indicators of the representation on which the algorithm is described. For example, for imperative programming languages, this is the complexity of the control flow, data flow, the size of the program and any of its elements. In fact, if there exists a certain coefficient Acof capable of considering these measurable characteristics of various implementations of the algorithm, then Expression (1) must be rewritten as follows

\[
(Acof \times cof \times C’/C) > 1
\]  (2)

Using Expression (2), we can give a definition of analyzability - Analyzability is a quantity Acof that distinguishes different implementations of the same algorithm, and the descriptions using the same presentation form, from each other that affects the time that a specialist needs to understand the principles of the algorithm.

IV. METHODOLOGY

The methodology of quantitative assessment could be implemented using two stages - recognizer construction described in Fig.1, and obtaining a quantitative assessment described in Fig. 2.

![Fig.1 Recognizer construction](image)

**Recognizer construction.** Carrying out all the necessary work on the collection, preparation of data, the development of a recognizer and training in its effective assessment of the Analyzability metric.
Obtaining a quantitative assessment (analyzability) on the basis of the recognizer obtained is the first step. At the stage of recognizer construction, the participation of experts is necessary to form an expert assessment and specialists to obtain reliable recognizer results. Once the recognizer model is obtained, it can be used to obtain a quantitative assessment - analysis without having the special knowledge needed in the first stage.

V. ANALYZABILITY MODEL

To develop the proposed software module, Fig. 3 demonstrated the main tasks that were solved.

JAVA programming language was selected for evaluating the program code of which a quantitative assessment will be calculated to measure the analyzability. The Development software is based on the work from [17][18], and A set of 41 software products have been collected from [19] and Fig. 4 shows a sample of the selected software product.

The selected software product was configured with a “reference analysis” that made up the training sample. A set of metrics were generated that is suitable for a given programming language and fully describe the source code of the programs, in so far as it relates to “analyzability”. The software metrics were extracted from [14] http://www.virtualmachinery.com/sidebar2.htm: The Halstead metrics and as described in table 1.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Metric Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>unique operators</td>
<td>(n1)</td>
<td>the number of distinct operators</td>
</tr>
<tr>
<td>unique operands</td>
<td>(n2)</td>
<td>the number of distinct operands</td>
</tr>
<tr>
<td>total operators</td>
<td>(N1)</td>
<td>the total occurrences of operators</td>
</tr>
<tr>
<td>total operands</td>
<td>(N2)</td>
<td>the total occurrences of operands</td>
</tr>
<tr>
<td>Program vocabulary</td>
<td>(n)</td>
<td>n = n1 + n2</td>
</tr>
<tr>
<td>Program length</td>
<td>(N)</td>
<td>N = N1 + N2</td>
</tr>
<tr>
<td>Calculated program length</td>
<td>(N')</td>
<td>N'=n1log2(n1)+n2log2(n2)</td>
</tr>
<tr>
<td>Volume</td>
<td>(V)</td>
<td>V = Nlog2(n)</td>
</tr>
<tr>
<td>Difficulty</td>
<td>(D)</td>
<td>D= (n1/2) * (N2/n2)</td>
</tr>
<tr>
<td>Effort</td>
<td>(E)</td>
<td>E= DV</td>
</tr>
<tr>
<td>The time required to program</td>
<td>(T)</td>
<td>T= E/18 seconds</td>
</tr>
<tr>
<td>Delivered bugs</td>
<td>(B)</td>
<td>B=V/3000</td>
</tr>
<tr>
<td>cyclomatic complexity</td>
<td>(Cc)</td>
<td>Cc= edges- Nodes + 2</td>
</tr>
</tbody>
</table>

A set of “confusing transformations” applicable to a given programming language has been formed. A study has been made of the influence of “confusing transformations” on the “analyzability” of the program code. A recognizer model was created using the Weka Tool. The topology, the number of hidden layers, neurons, the activation function and the learning algorithm of an artificial neural network was selected. The smart program was written in C# to extract the software metrics described in Table 2, all forty-one (41) program codes were classified as ‘analyzability’ with the help of software engineering experts. Then, a group of seven professional programmers were required to modify the forty-one (41) program codes to complicate the control flow, data flow, the size of the program and any of its elements. This collection is called “confusing transformation” and considered as “reference unanalyzable” software. The modified program codes with the support of the experts were classified as non-analyzability. Finally, the data set was processed, developed and prepared for the classification process. Fig. 5 describes a sample of the developed data set. Each code example is designed as a separate application and represents a full-fledged Java software. Since each of these applications was prepared by the company’s specialists as a “reference-analyzable” example of using the Java language, an assumption is made about the “reference-analyzability” analysis of these applications as one of the indicators of their quality.
For a metric evaluation of software products, a set of thirteen (13) metrics were generated that quantitatively measure various properties of software.

Fig. 5 Data set sample

The proximity matrix was setup and calculated to discover more information about the dataset. Fig. 6 shows the proximity matrix analysis process.

Fig. 6 A sample of Proximity Matrix

Using “Weka” classifiers functions, a multilayer perceptron was selected for the classification process and to generate the analyzability model. Since at the moment, there are no reliable methods for the analytical determination of the number of neurons, hidden layers, activation function, and method of training Multilayer-Perceptron, the selection of the best configuration was carried out by completely sorting out the possible values for these characteristics. With the help of Weka 3.9 tools. Fig. 7 and Fig. 8 show a sample of the stages of choosing the best Multilayer-Perceptron configuration, which shows the configuration of the training and the ability to generalize Multilayer-Perceptron with one hidden layer and ten neurons in the hidden layer.

Fig. 7 Sample of Multilayer-Perceptron configuration

This work also calculates the information-gain for each software metric selected in this work, and Fig. 9 demonstrates the results. The information gain was carried out to identify the metrics that have the greatest influence on analyzability.

Fig. 8 A sample of the generated Multilayer-Perceptron layers

It can be seen from the diagram that the indicated metrics are more influenced by the assessment and analyzability. The finding also showed that Efforts, Time and Bugs play an important role in defining the level of analyzability. The main purpose of the experiment was to prove the effectiveness of the proposed approach to assessing the quality indicator “analysability”. The following can be distinguished as additional objectives of the experiment: Specialists in the field of software development, with a total number of seven software engineers who worked as experts to evaluate and test the software in terms of the quality of “analyzability”. Before the start of the experiment, we explained the essence of the experiment and proposed to evaluate each of the software products in terms of the quality of “analyzability”. After that, the values obtained on the basis of the methodology of quantitative assessment - analyzability were compared with an expert assessment.

In order to evaluate the diagnostic effectiveness of the method, considering the consequences of false decisions, characteristic curves were used. These reflect the mutual dependence of false-positive and true-positive results. To quantify the methods used, a comparative analysis of the areas under the ROC curves was implemented.
The curve of the proposed method has a larger area than the curve of the expert's evaluation. As can be seen in Fig. 10, the values obtained using the quantitative assessment methodology - analyzability, and Fig. 11 shows the values of “analyzability” obtained using expert evaluation are quite well connected. This indicates the effectiveness of the developed methodology in assessing the “analyzability” of software products.

The ROC analysis method allows us to determine the quantitative value of the reliability of the differences in the information content of the studied methods. To do this, we calculated the area under the curves (Fig. 9.5) and, using special formulas, establish the confidence interval in the difference in the information content of the methods. It is generally accepted that the coefficient of the area of the curve in the range of 0.9-1.0 should be considered as an indicator of the highest informative of the diagnostic method - 0.8-0.9 - good, 0.7-0.8 - acceptable, 0.6 - 0.7 - weak, 0.5-0.6 - extremely weak.

In Fig. 10, ROC-curves were constructed to compare the information content of the proposed method. The Curve is closer to the upper left corner with ROC=7.4. Therefore, it is more informative than a curve of expert evaluation with ROC=7.3. Both methods are accepted. However, the proposed method outperforms the expert evaluation. This experiment confirmed that “analysis” is not described by anyone metric but is a complex combination of several metrics. It is therefore important to note that the study made it possible to obtain a number of new scientific and technical results that are essential for quantitative assessment of the quality of software products. Based on the analysis of the state of the issue, in the field of evaluating and controlling the “analyzability” of software products, the main shortcomings of existing approaches were identified. This analysis made it clear of the need for a formalized definition of “analyzability”.

A methodology was developed for evaluating the program code for the subject of “analyzability”, which allows one to obtain an objective quantitative assessment of the quality indicator “analyzability” in accordance with the requirements, limitations, purpose and specific features of the product. For validity, a program code recognizer was developed according to the “analyzability” quality indicator, with the ability to recognize a “new” program code, i.e. with the ability to generalize. It was found that for test software from thirteen (13) measured metrics, 98% of “analyzability” represents only seven software code metrics, the remaining six metrics make up only ~ 5% of “analyzability”. Similarly, a set of “confusing transformations” was formed using Java programming language, which affects the “pars ability” of program code. In the end, a study of the influence of “confusing transformations” on the metrics of the program code was carried out to allow us to create tools for studying the properties of program codes and its use in the construction of a recognizer - analyzer. The analytical representation of the “analyzability” of the program code makes it possible, when designing software products, to set quantitative characteristics of product properties. For reliability, a complex of algorithms was then developed to implement an automated solution to the problem of a quantitative assessment of the quality indicator “analyzability”. On this basis, a software complex was equally used to implement a quantitative assessment of analyzability through a dialogue with the user.
The resultant effects showed that the conducted simulation reflected a satisfactory convergence of the results of the expert assessment of “analyzability” and —analysis, measured by the developed methodology.

V. CONCLUSIONS

In conclusion, the experiments performed in this study discovered three major metrics (Effort, Time and Bugs) that potentially define the analyzability level as explained earlier. Since the aim of the proposed research work is to develop a methodology for quantitative assessment considering the analyzability quality indicators of software products, it is pertinent that solutions are provided based on the analysis in the field of evaluation and control of the software products, where the main weaknesses of existing approaches were earlier identified. The scientific novelty is that the problem was formulated and set-out for obtaining a formalized assessment of the quality indicator “analyzability”, in accordance with the concept of analyzability of program code as introduced, defined and formalized.

During the methodology phase, a quantitative assessment of the quality indicator “analyzability”, a recognizer analyzer were developed that was capable of effectively evaluating the “analyzability” of program codes. A set of metrics was identified that allows us to essentially evaluate the quality indicator “analyzability” in a particular class of programming languages. An analytical representation was obtained - of the program code being analyzed, which confirms the effectiveness of the proposed “analysis” assessment. The practical value of the work showed clearly the required quantitative value of the “analyzability” quality indicator as a technical task for software development. This helps maintain this value throughout the entire software product development cycle and helps in making decisions based on quantitative indicators of a high-quality product in terms of “analyzability”. At the conclusion of the study, a practical tool was also created for assessing the quality of software products in terms of “analyzability” to generate a competitive environment in order to form a set of metrics for the program code, and completely evaluate the quality indicator “analyzability”. This provided an avenue to round up by formulating a confusing transformations test directly to affect the analyzability of the program code in order to test the influence on the program code metrics. The results ultimately produced an analytical representation from the program code which eventually confirms the effectiveness of the proposed “analyzability” assessment in this study. Finally, the outcomes of this study represented a set of algorithms that could implement an automated solution to solving the problems of assessing the software quality indicator “analyzability”.

REFERENCES

9. . Vol.6, No.3.
21. n:6:4.9

AUTHORS PROFILE

Dr. Boumedyen Shannaq, is a skilled university professor for 10 years, primarily teaching Information Systems and Information Technology Courses. Interested in Electronic Learning, Innovation, Artificial Intelligence, Analysing Big Data. Professional skills in Public Speaking, Time Management, Self-motivation, Inter-personal communication and Record-keeping. Focused on Analyze ideas and use logic to determine their strengths and weaknesses. Manage oneself, people, time, and things. Obtained a Ph.D. in Information Systems (data engineering-Technical Issues).

Dr. Richmond S. Adebayo, is a Faculty at the University of South Carolina Upstate. He is CISSP & CISM certified. He has over fourteen (14) years’ experience teaching and developing various courses in Information Systems, Digital Forensics and Security. He also has over 22 published research articles in various high impact scholarly journals. His main expertise and research interests involve Security Architecture, Data Analytical Science, Privacy-enhancing technologies impression, Single sign-on, Security Level Management, Audit-proof Network management and protocols.

Published By: Blue Eyes Intelligence Engineering & Sciences Publication

Retrieval Number: F7780038620/2020/6BEIESP
DOI:10.35940/ijrte.F7780.038620

2408