DEVELOPMENT OF A HYBRID METHOD FOR CALCULATION OF SOFTWARE COMPLEXITY

T.H. KAZIMOV, T.A. BAYRAMOVA

Abstract. The use of code metrics allows software developers and project managers to evaluate various features of the software (to be built or already in existence), predict workload, determine software complexity and reliability, and quantify the quality of software systems being developed. Articles written in recent years have proposed various methods for solving this problem. However, there is still no very effective approach to measuring software complexity. This article provides a brief overview of existing software complexity metrics and proposes a new hybrid method for computing software complexity. The proposed hybrid method for evaluating software complexity combines the key features of the Halsted, McCabe, and SLOC metrics and also allows for a more efficient assessment of complexity.

Keywords: software engineering, software complexity, complexity metrics, hybrid method.

INTRODUCTION

One of the most dynamically developing spheres of modern life is information technology. Today, people’s lives are organized in such a way that even people who are far from information technology use them to achieve their goals. The introduction of automated information systems not only reduces the number of operations performed by a person, but also creates new problems. As software grows in size and complexity, the number of bugs increases. Software projects that at first glance appear to be successful may be stopped due to such errors, or the code may be rewritten. This leads to a slight increase in budget expenditures. The world’s leading software companies are working on these problems. Various standards organizations (ISO/IEC, IEEE) have developed hundreds of standards covering all stages of software life to improve its quality. Software quality is a complex and multifaceted concept. After the advent of computers in the defense industry in the 1970s, metrics began to emerge to evaluate software performance and quality management.

Software measurement is one of the most important issues in developing quality software. In the words of Tom DeMarco [1], “you can’t control what you can’t measure”.

At the moment, a large number of different software metrics have been developed. Software metrics fall into three categories (Fig.1):
Development of a hybrid method for calculation of software complexity

- **Product metrics** are product characteristics such as performance, design, size, quality level, and complexity.
- **Process Metrics.** These metrics are used to improve software development and maintenance processes.
- **Project Metrics.** These metrics describe the performance and characteristics of the project.

Among these indicators, software complexity indicators are of particular importance. In the 10th edition of Sommerville’s book Software Engineering, he noted that one of the most important tasks in the development of the modern software industry is the management of software complexity [4]. Research shows that when the project size approaches 5 million lines, the number of defects begins to increase dramatically (Fig. 2). This can be explained by the significant complication of large-scale projects [5].

![Software metrics](image1)

**Fig. 1.** Software metrics

![Dependence of KLOC](image2)

**Fig. 2.** Dependence of KLOC

Complexity is the degree where it is difficult to verify and understand the design or application of a system or component. The problem of software complexity estimation is widely studied in the field of software engineering. The complexity of the software may vary depending on the choice of algorithm, design, choice of programming language, writing code. Research shows that increasing complexity increases the number of bugs in software, which makes it difficult (in some cases impossible) to maintain and improve programs and makes it difficult to test certain modules. The legibility of the program code directly depends on the level of complexity of the program. The relationship between indicators of software complexity and various attributes of a software system is shown in Fig. 3.

The importance of program complexity indicators can be illustrated as follows [7]:

- Difficulty metrics can help people predict and sustain projects.
- Complexity metrics can help estimate the amount of programming and development costs, and estimate maintenance costs.
Difficulty scores can help you choose the most suitable program with the same functionality. The less complex the programs, the better they are.

Complexity measures can be used to predict defects or errors [8].

By determining the complexity of software systems, the overall workload and personal contribution of each person working on software modules to the overall work can be assessed in evaluating the performance of software developers [9].

**RELATED WORK**

Application of program code metrics allows professionals who work on the project to evaluate various features of existing or to be created software, to predict the scope of work, quantitatively characterize these or other project solutions, to evaluate quality of prepared systems, complexity and reliability of software [9–13].

The large amount of data and functionality required in today’s enterprise systems presents many challenges for software developers. It is difficult to maintain a balance between software complexity and ease of use, which are indicators of the quality of software, as the complexity of the software increases and its use becomes more difficult. To ensure the quality of the software and the high level of project management, it is necessary to control the complexity and other related characteristics. Large companies such as SAP and Oracle are currently losing market share due to the cost and complexity of the software product [14].

Numerous studies show that complexity leads to an increase in the number of vulnerabilities in software. Programmers face a number of challenges when trying to make changes to complex software components. Software developers should strive for minimal complexity, as increasing complexity creates security risks that not only damage the business, but also damage its reputation. Violation of safety measures can also pose a serious threat to human health. Some experts deliberately complicate the same program so that they do not write it twice. It is necessary to simplify the management of software security. The authors substantiate the relationship between security and software complexity (Fig. 4) [15].

Different software development organizations use different metrics to measure and maintain the quality of software code. In [16, 17], the authors...
Development of a hybrid method for calculation of software complexity conducted comparative analytical studies using Halstead metrics and proved that software complexity metrics can be used to measure various characteristics. The authors analyzed the program code written for the same function in the programming languages Python, C, JavaScript, and Java. The complexity of these program codes was calculated based on the Halstead scores. Experience has shown that python is more convenient and simpler, while Java is a modern, powerful, but complex programming language. They showed that program performance is important throughout the life of a project, and they concluded that early calculations have a direct impact on fixing errors and thus saving time and budget.

In [18], the authors point out that the main reason for software vulnerabilities is its complexity. Failure to discover software vulnerabilities in a timely manner can jeopardize confidentiality, completeness, and availability. Studies of 12 programs, differing in characteristics and sizes, have shown that the complexity of all software components (size, structure, etc.) is important for predicting vulnerabilities.

In their articles, Shin and William examined code complexity metrics that can be used to predict software vulnerabilities, and concluded that the more complex the software code, the more vulnerabilities. Their results showed a correlation between difficulty scores and vulnerabilities in the Mozilla JavaScript Engine [19].

In [20], the authors investigated the influence of complexity, coupling and cohesion on the number of vulnerabilities in a program. Experiments with Mozilla Firefox have empirically proven that there is a relationship between these parameters and vulnerabilities in the program. Vulnerabilities cannot be discovered in the early stages of software development. Metrics such as complexity, coupling and cohesion can be calculated early. Calculating these metrics can help software developers identify potential vulnerabilities.

In [21], the authors presented a new approach to measuring the quality of software using fuzzy metrics obtained as a result of software design. This metric begins with an assessment of the difficulty rating for each class, which is itself assessed based on the difficulty rating of the class attributes and the difficulty rating of the class methods. Then, to assess the quality metrics of the fuzzy code,
they followed a pattern to explain the relationship between class complexity and LOC for that class, and between method complexity and LOC for that method.

The paper [23] examines the problem of software complexity and its impact on software errors. Having studied various studies, the authors came to the conclusion that with the increase in the complexity of the software system, the threats to its security also increase. While most studies have shown that bugs in software systems are related to their complexity, some researchers have noted that the relationship between complexity and the occurrence of bugs is weak [24].

The software development process, including documentation, design, software, testing, and support, can be measured statistically. Thus, the quality of the software can be effectively controlled. Software metrics are very important in software engineering research [25] provides a summary of software metrics and their types. The authors note that measuring the complexity of software is an integral part of software performance and affects the price and reliability of software products.

Software users assume there are no bugs in the program, but software developers know that it is very difficult, and in some cases impossible, to write program code without bugs. This complexity is mainly due to the intrinsic complexity of the program and the problems that arise when developing and testing the program. In [26], the authors demonstrated the relationship between the complexity of software and its reliability in each specific case.

In [27], the authors analyzed aspects influencing the complexity of the program based on various metrics. A large amount of program code to some extent affects the programmer’s thought process and leads to more errors in the program. At the same time, if the number of modules and branches in the program is large, the volume of information exchange is large, and the program is very complex, this will cause problems for the program testers and reduce the quality of testing.

Software complexity plays an important role in reducing the effort required to build and maintain software, and in improving testing efficiency and software quality. The more complex the solution of the program, the more errors it creates [28] examines four software metrics, their importance, strengths and weaknesses. The authors come to the conclusion that each method covers a part and takes into account a certain group of parameters. Therefore, it is important to use a combination of these metrics to measure software complexity.

HYBRID SOFTWARE COMPLEXITY CALCULATION METHOD

The hybrid software complexity method presented in this study combines the advantages of the SLOC, Halstead, Maccab metrics and the module connectivity metric.

**Number of lines of code (LOC).** When assessing the complexity of a software product, three groups of indicators are usually used: indicators that determine the size of the program, indicators of the complexity of the program flow, and complexity of the data flow of the program. The first group is more common because the metrics are simple. The program size traditionally means the number of source lines of the program (SLOC – Source Lines Of Code) [29].
Lines of Code (LOC) or Source Lines of Code (SLOC) are used to measure the size of a program by counting the number of lines in the text of the program’s source code. LOC measures the amount of code, can be used to compare or evaluate programs that use the same programming language and are coded using the same coding standards. Lines of code are completely dependent on the programming language and may differ during the conversion of the program code to any other programming language. This indicator cannot be considered effective in assessing complexity.

**Halsted metric.** M.H. Halsted was the first researcher to write a mathematical formulation of software metrics [30]. Table 1 shows the main Halsted metrics.

<table>
<thead>
<tr>
<th>Table 1. Halsted metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program dictionary</td>
</tr>
<tr>
<td>Program length</td>
</tr>
<tr>
<td>Program scope</td>
</tr>
<tr>
<td>Program complexity</td>
</tr>
<tr>
<td>Implementation effort or program comprehensibility</td>
</tr>
<tr>
<td>Expected number of errors in the program</td>
</tr>
</tbody>
</table>

The Halsted metric is based on four measurable program characteristics:

- $h_1$ — the number of unique program operators;
- $h_2$ — the number of unique software operands;
- $N_1$ — total number of program operators;
- $N_2$ — total number of program operands.

**McCabe’s cyclomatic complexity.** McCabe proposed this metric in 1976. McCabe’s cyclomatic complexity metric is one of the most common indicators for evaluating software complexity. This indicator is calculated based on the control-flow graph of the program and is an indicator of the complexity of the control structures.

The formula for calculating cyclomatic complexity is as follows:

$$C = e - n + 2p,$$

where $e$ — the number of edges of the graph; $n$ — the number of nodes of the graph; $p$ — the number of connected components.

This method takes into account all kinds of cyclic and conditional operators, as well as the complexity of their logical predicates. Linear operators are ignored here. Despite the passage of 45 years, it remains relevant because it more accurately expresses the complexity of the fundamental structures of software. The author of the metric notes the high correlation of the metric with errors. Therefore, the use of this metric to evaluate errors is reasonable and logical. Cyclomatic complexity allows not only to estimate the labor costs and the cost of software projects, but also to make the necessary management decisions taking into account risks [32, 33].
The program module coupling metric distinguishes between data communication, data structure, control, common area (global data), content, as well as external communication, message communication, subclass communication, time communication, and no communication. The advantages of these metrics are that they were developed back in the 80s, are well studied, are often used, and the calculation of values for metrics is automated. These metrics are intended for structured programming, but can be applied to object-oriented programming.

A hybrid method for calculating the complexity of a program code. The proposed method provides an integrated approach to the complexity of the software system and takes into account the following disadvantages of the McCabe and Halsted metrics and LOC:

- it cannot be argued that a program with the maximum number of lines of code is more complex, sometimes a small program can be quite complex;
- linear, cyclic, conditional operators in the system of Halsted metrics have the same complexity, which is not true;
- all cyclic, conditional operators in the McCabe metrics system have the same complexity, which is not true. It is assumed that the more nested cyclic and conditional (multiple branching) statements, the more complex the program code becomes and the greater the likelihood that the programmer can make a logical error;
- models do not take into account the inter-module structural complexity of the software, which generates a significant number of defects.

In this method, the main parameters affecting the complexity are the algorithmic complexity of each program module, the number of standard called programs ready for use, and the total number of intermodule links.

When solving problems algorithmically, it often becomes necessary to create a cycle containing another cycle in its body. Such loops are called nested loops. Sometimes you have to check several conditions in a row. To solve the problem of redundancy, you can nest conditional statements inside each other. This is called multiple branching of conditional statements. Nested loops and multiple branches in the code complicate it. At the same time, the number of interacting modules and standard software applications available in these modules directly affects the complexity of the program.

The complexity of the program code can be calculated using the formula (higher value is considered better):

\[ C = \log_2 E + \log_2 (N - n) + r , \]

where \( N \) — the number of all operators in the program code; \( n \) — total number of conditional and cyclic operators; \( r \) — the number of intermodular couplings; \( E \) — parameter depending on the number of cyclic and conditional operators in the program code. This parameter is calculated using the following formula:

\[ E = \sum_{i=0}^{n} (i+1)m_i , \quad (i = 0,1,\ldots,n) , \]
Development of a hybrid method for calculation of software complexity

$m_i$ is the total number of conditional and cyclic operators in which conditional and cyclic statements of number $i$ are nested. (For example, if there are 2 cyclic statements in the program code and three nested cyclic statements in each of them, then $i=3$, $m_3=2$).

Let's calculate the complexity of two programs written in C++ (Fig. 5 and 6). First, we received an expert assessment of the complexity of these programs based on the assessment of 3 experts on a 10-point scale. All experts rated the algorithm in Fig. 6 as complex, since this algorithm has more nested cyclic and conditional operators. The calculation results are shown in Table 2.

```cpp
#include<iostream>
using namespace std;

int main()
{
    int n=3;
    int a[n][n];
    int m=0
    for {int i=1; i<=n; i++}
        for {int j=1; j<=n; j++}
        {
            cin>>a[i][j];
            if (a[i][j]==i+j)
                m=m+1;
            else
                m=m;
        }
    if (m>0)
        for {int i=1; i<=n; i++}
            for {int j=1; j<=n; j++}
            {
                if (a[i][j]==i+j)
                    cout<<a[i][j];
                else
                    continue;
            }
    else
        cout<<"No such element";
    return 0;
}
```

**Fig. 5.** Finding complexity of program 1

<table>
<thead>
<tr>
<th>Table 2. Software complexity metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complexity metrics</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>LOC</td>
</tr>
<tr>
<td>Expert assessment</td>
</tr>
<tr>
<td>MacCabe</td>
</tr>
</tbody>
</table>

Системні дослідження та інформаційні технології, 2022, № 2
Continued Tabl. 2

<table>
<thead>
<tr>
<th>Complexity metrics</th>
<th>Calculation</th>
<th>Program 1</th>
<th>Program 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hybrid</td>
<td>$E = \sum_{i=0}^{n} (i+1)m_i$</td>
<td>$N=108, \ \ n=5, \ r=0, \ m_0=3 \ (i=0), \ m_1=2 \ (i=1)$</td>
<td>$N=101, \ n=5, \ r=0, \ m_0=3 \ (i=0), \ m_1=1 \ (i=1), \ m_2=1 \ (i=2), \ m_3=1 \ (i=3)$</td>
</tr>
<tr>
<td></td>
<td>$C=\log_2 E + \log_2 (N-n)+r$</td>
<td>9,49</td>
<td>9,58</td>
</tr>
</tbody>
</table>

As can be seen from Table 2 according to McCabe, the complexity of programs is the same, because in this method all conditional and cyclic operators have the same level of complexity. The hybrid method takes into account the nesting of cycles, so the result is closer to the expert assessment.

In another example, let’s calculate the complexity of a program code written for the same function in C++ and Java using the metrics mentioned above (Fig. 7 and 8). Depending on the chosen programming language, the complexity may increase. Calculating the complexity will help you choose the simpler of the two programs. The calculation results are shown in Table 3.
Public Class
{
    Public static void main (String[] args)
    {
        int arr[10]={1,9,0,5,6,7,8,2,4,3};
        int length = 10;
        int result[]=new int (10);
        result=QuickSort (arr,0,length-1);
    }

    Static int [] QuickSort (int []a,int r)
    {
        if (l<r)
        {
            int i=l;
            int j=r;
            int k = (int)((l+r)/2);
            int pilot = a[k];
            do
            {
                while (a[i].less (pilot))
                {
                    i++;
                }
                while (pilot.less (a[i]))
                {
                    j--;
                }
                if (i<=j)
                {
                    int t = a[i];
                    a[i] = a[j];
                    a[j] = t;
                    i++;
                    j++;
                }
                while (i<j)
            }
        }
        return a;
    }

    } // end of class QS

#include<iostream>
using namespace std;
int main()
{
    int A[]={1,9,0,5,6,7,8,2,3,4}
    int length = 10;
    quickSort(A,0,length-1);

    void quicksort(int A[],int F,int L)
    {
        int pivotIndex;
        if (F<L)
        {
            Partition(A,F,L,pivotIndex);
            quicksort(A,F,pivotIndex-1);
            quicksort(A,pivotIndex+1,L);
        }
    }

    void partition(int A[],int F,int L,int & pivotindex)
    {
        int piv = A[F];
        int lastS1 = F;
        int firstUnknown = F+1;
        for (firstUnknown = F+1; ++firstUnknown)
        {
            if (A[firstUnknown]<pivot)
            {
                Swap(A[firstUnknown],A[lastS1]);
                Swap(A[F],A[firstUnknown]);
                lastS1 = ++lastS1;
            }
        }
        return A;
    }

    void Swap(int & x,int & y)
    {
        int temp = x;
        x=y;
        y=temp;
    }

Fig. 7. Quick sort implementation code in C++ Fig. 8. Quick sort implementation code in Java
Table 3. Software complexity of C++ and Java programs

<table>
<thead>
<tr>
<th>Complexity metrics</th>
<th>Calculation</th>
<th>C++</th>
<th>Java</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOC</td>
<td>Number of lines of code</td>
<td>48</td>
<td>52</td>
</tr>
<tr>
<td>MacCabe</td>
<td>For a single program ( C = e - n + 2 ) ( e=28, n=26 )</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>( N_1=142 ) ( N_2=57 ) ( h_1=25 ) ( h_2=21 )</td>
<td>33,92</td>
<td>39,75</td>
</tr>
<tr>
<td>Halstead</td>
<td>( D = \left( \frac{h_1}{2} \right) \ast \left( \frac{N_2}{h_2} \right) )</td>
<td>( N=142, n=3, r=0 ), ( m_1=1 ) (i=1), ( m_0=2 ) (i=0)</td>
<td>9,11</td>
</tr>
<tr>
<td>Hybrid</td>
<td>( E = \sum_{i=0}^{n} (i+1)m_i ) ( N=133, n=5, r=0, m_0=3 ) (i=0), ( m_1=1 ) (i=1), ( m_2=1 ) (i=2)</td>
<td>33,92</td>
<td>39,75</td>
</tr>
</tbody>
</table>

CONCLUSION

Software complexity metrics are one of the key aspects of software process management. Complexity makes software difficult to understand, which creates problems when maintaining a software system and adding new functions to it. In the articles of recent years, various methods of solving this problem have been proposed. Numerous studies in the field of software complexity metrics suggest that there is no universal metric for assessing the complexity of any program code. The use of any metric, hybrid metric, or multiple metrics depends on the specific problem. This article proposes a new hybrid method for calculating the complexity of software, the effectiveness of which is substantiated by experiments. The proposed metric makes it possible to clarify (due to additional indicators) the complexity of software modules of a large class separately from the known metrics. Based on the above, it is concluded that software metrics are important at the stages of the project life cycle and, with early application of program metrics, help to largely overcome the presence of errors and, thus, save time and money.

REFERENCES


INFORMATION ON THE ARTICLE

**Tofig Kazimov (Hasanaga),** ORCID: 0000-0001-9245-6731, Institute of Information Technology of Azerbaijan National Academy of Sciences, Azerbaijan, e-mail: tofig@mail.ru

**Tamilla Bayramova (Adil),** ORCID: 0000-0002-8377-3572, Institute of Information Technology of Azerbaijan National Academy of Sciences, Azerbaijan, e-mail: toma_b66@mail.ru

**РОЗРОБЛЕННЯ ГІБРИДНОГО МЕТОДУ ОБЧИСЛЕННЯ СКЛАДНОСТІ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ / T.G. Kazimov, T.A. Bayramova**

**Анотація.** Застосування показників програмного коду дозволяє розробникам програмного забезпечення вибирати керівників проектів оцінювати різні функції програмного забезпечення, яке буде створено або вже існує, прогнозувати робоче навантаження, визначати складність і надійність програмного забезпечення і надавати кількісні характеристики якості розроблюваних програмних систем.

У цих останніх роках пропонуються різні методи вирішення цієї проблеми. Утім, до сьогодення немає ефективного підходу до вимірювання складності програмного забезпечення. Подано стислий огляд існуючих метрик складності програмного забезпечення. Запропоновано новий гібридний метод обчислення складності програмного забезпечення, який об’єднує ключові характеристики метрик Холдстеда, Маккабі i SLOC, а також дозволяє більш ефективно оцінювати складність.

**Ключові слова:** програмна інженерія, складність програмного забезпечення, метрики складності, гібридний метод.