Complexity Assessment based on UML-Activity Diagram

Maushumi Lahon, Uzzal Sharma

Abstract: Assessing complexity can significantly contribute to the attainment of the various quality attributes associated with a system. The avoidable complexity can be identified and reduced on the basis of the assessment. It holds the key to success of the system being developed. Various evaluation methods exist which have specific objectives and basis and all contribute to enhance product quality. In this paper a Complexity Assessment approach based on Activity Diagrams (CAAD) is proposed to evaluate the process view of the architecture of a system. The proposed approach estimates the complexity of the system/class/function from the UML representation of the process view of the architecture in the form of activity diagrams. This complexity measure may be used to assess and estimate the time and effort required to develop the system. This approach can estimate the coding complexity in terms of size without actually developing the code for the system/class/function. The paper is on calculating a complexity factor C from the given activity diagram and further develop a relationship between C and LOC metrics.

Index Terms: Software development, Activity Diagrams, Activity flow graph, Complexity, Metrics

I. INTRODUCTION

Complexity is an area related to every domain of work. In the area of software development, complexity has great impact on the development effort and cost. So to discuss the challenges of this area, research works are being carried out at different levels with different measures. It becomes difficult to achieve the desired quality of software as they are growing in size and thus increasing the complexity of developing them [1].

- This work presents an approach to evaluate the architecture of a software system from the process view architectural representation of the system. The approach is based on 4 + 1 view representation of architecture proposed by Kruchten in his work [2].
- The objective of the approach is to evaluate the architecture by calculating the complexity of the system from the process view representation of the architecture by using metrics to calculate complexity. Unified Modelling Language (UML) is used, as it is considered as industry standard modelling language with a rich graphical notation, and comprehensive set of diagrams and elements. Both the structural and behavioural aspects of the system can be represented using the various UML diagrams.
- The proposed approach “Complexity Assessment based on Activity Diagrams” (CAAD) evaluates the system/class/function to determine the development complexity. This could be used as an indicator for estimating the effort in implementing the other quality attributes like modifiability, cost and benefit, maintainability used in other approaches for evaluating the architecture.
- The focus of this work is to find a measure of complexity of the system/class/function in terms of developing it. This measure can facilitate the process of estimating the effort and time required to develop the system without actually developing the code. It is an essential part of system development to estimate the size of the system which introduces different types of complexity in developing the system. This complexity could be in terms of effort and time which has to be converted to person-month estimate. This finally can contribute to the cost estimation of developing the system.
- The paper proposes the estimation of a complexity measure C from the activity diagram representation. This activity diagram could be of a system, a class or a function. The activity diagram is converted to an activity flow graph from which the metrics for calculating C is derived. Higher C value indicates more development complexity.
- Finally, an implementation of the approach on a project was made to develop a relationship between the C value and LOC metrics to establish that higher C values do indicate more complexity.

The rest of the sections are organised as follows: Section II covers the review of the previous works on metrics and complexity, Section III gives the methodology of calculating the complexity value from the activity diagrams along with a case study. Section IV highlights the various points with respect to the proposed complexity measure approach in comparison with complexity measures derived from other UML diagrams, section V presents the implementation of the proposed approach on a project to calculate the complexity value from different activity diagrams used in the selected project. Section VI presents the Analysis and Results from the implementation and finally section VII draws the conclusion and suggests directions for future work.

II. PREVIOUS WORK

The term complexity is a familiar term and forms a part of almost every domain. More functionality induces more complexity, which in turn is related to effort and cost. Moreover, complexity is directly related to quality aspect, more requirements in quality factors introduce more...
complexity in the development phase. Researchers working in this area of complexity in different domains have developed metrics and measures to assess and deal with the context. In the software engineering domain also design complexity and code complexity is a significant area of concern and is a matter of interest even today. Weyukar's work on evaluating software complexity measures suggest that complexity measures was an area of interest during that period and software complexity measures already existed then which were evaluated in the work [3]. Zuse in his book has described and analysed different software complexity measures and methods and states that complexity is a multidimensional property which cannot be estimated by a single magic number [4]. Review of complexity metrics [5], use of complexity metrics in for analysis in different areas [6] and design of different complexity metrics [7,8,9] clearly establish the fact that, study in this area is of much interest to the research community. A work on assessing architecture complexity way back in 1998, proposed the Interactive Architecture Pattern Recognition (IAPR) system for locating user-specified at varying levels of abstraction patterns within software architecture. No concept of metrics was involved in this [10]. Software Architecture Complexity Model (SACM) was proposed based on theories from cognitive science and system attributes that have proven to be indicators of maintainability in practice [11]. The model emphasise difficulties in understanding software architecture and the elements that complicate the verification of implemented and intended architecture. Tong Yi et.al in their work on comparisons of metrics for UML class diagrams represented a comparison based on different aspects [12]. The author considers six existing metrics to measure the complexity for UML class diagrams and compares the metrics on the basis of different viewpoints. Kang et.al uses weighted class dependence graphs to represent a class diagram. The structural complexity of the class diagram is calculated based on entropy distance which is an objective measure [13, 14]. Marchesi’s proposed metrics focused on use-case diagrams and class diagrams which intends to measure the orientation of the object and quality of the system. The proposed metrics aims to measure complexity, responsibility balancing among packages and classes, as well as cohesion and coupling among the entities of the system [15]. Sengupta and Kanjilal proposed the Component Architecture Complexity Measurement Metrics (CACMM) on UML component diagrams. They used Component Architecture Graph for graphically representing the component diagram. This graphical representation was analysed to measure complexity at individual component level, component to component level and the overall architecture. The metrics can be used to redesign to minimise coupling and create highly cohesive components [16]. Complexity of a system can be assessed at two different levels, the architecture level and the behaviour level [17]. At the architecture level both formal and informal notations are used for assessment, but most often complexity is assessed at the code level applying the industry standard code metrics as McCabe Complexity metrics [18], Halstead Complexity Metrics [19] and Zage Complexity Metrics [20]. Other approaches to assess complexity are based on calculating the program elements like data types, number of methods and arguments of a function. Mesa et.al [21] in their work on complexity assessment measure has based their assessment measure on the tasks required for assembling and disassembling. The objective of the work is to measure complexity so that it can be reduced to give significant impact on the lifecycle of the system. The importance of complexity measure in the early stages have been seen in other works stating that complexity measure is an early indicator of the effort required for testing, implementing quality factors and minimising cost [22,23,24]. The presence of subjective component in complexity analysis is also considered to be a significant aspect which requires to be addressed while measuring system complexity [25]. The proposed approach considers the activity diagram from UML diagrams to capture the process view of the architecture. Literature reveals that there are different variations and representations of the UML notations depending on the purpose and research objective. As basic UML diagrams do not represent the artifact necessary for different types of evaluation, the notations are tailored according to the requirements of research. Such representation is seen in dependability evaluation of an architecture using Dependability Modelling and Analysis (DAM) annotations [26]. Similar work on reliability evaluation is found using DAM profile [27]. In connection to activity diagrams Petri net representation of UML 2 activity diagrams and Coloured Petri Net (CPN) representations are found in the literature review. CPN representation was used for execution validation [28]. The proposed complexity assessment approach CAAD is based on UML activity diagram to represent the system complexity from the dynamic aspect using a simple node and edge concept without going into the complex representations like Petri nets or CPN.

III. COMPARISON OF THE PROPOSED APPROACH WITH OTHER COMPLEXITY MEASUREMENT METHODS USING UML DIAGRAMS

There are different methods proposed to measure complexity of a system using UML diagrams. As no relevant work on measuring system/class/function complexity based on activity diagram was found in the review, a comparison of the proposed approach is made with complexity assessment approaches based on use-case diagrams and class diagrams proposed by Marchesi [15].

In case of complexity assessment based on Use-case diagram, four metrics were proposed –

- Number of Use Cases
- The overall number of communications among use cases and actors
- The total number of communications among use cases and actors, without redundancies introduced by extend and use relationships.
- Estimate of the global complexity of the system (assumption is that the key complexity measurement is the number of independent communications, metric UC3).

From the proposed metrics it is clear that the estimate of complexity is based on what is to be done by the system, not how. Complexity is introduced when the how component is decided as the resources required can be estimated. This estimate covers the different scenarios but not the dynamic aspect of the system.
In case of complexity assessment from class diagrams, the following metrics are not only useful to estimate complexity of the system, but may also be used to get a view of its quality. The metrics includes:

- The weighted number of responsibilities of a class both inherited and not inherited.
- The weighted number of dependencies
- Depth of inheritance tree

Number of immediate sub classes of a class

There are many assumptions in this complexity measuring method and the constants used are based on subjective judgement.

The following table summarises the important points of the three complexity assessment approaches based on three different UML diagrams

<table>
<thead>
<tr>
<th>UML diagram used for complexity assessment</th>
<th>Objective</th>
<th>Parameters considered</th>
<th>Diagram Type</th>
<th>Assessment Type</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use Case diagram</td>
<td>Early estimate of complexity of the system</td>
<td>Use cases, actors and communications</td>
<td>Static</td>
<td>Uses subjective assessment</td>
<td>1: Involves simple mathematical calculations. 2: Do not consider testing aspect in complexity measurement. 3: Covers Use case view.</td>
</tr>
<tr>
<td>Class diagram</td>
<td>Measure the expected complexity and quality of single class</td>
<td>Class responsibilities, dependencies, number of subclasses</td>
<td>Static</td>
<td>No subjective assessment</td>
<td>1: Involves complex mathematical calculations. 2: Do not consider testing aspect in complexity measurement. 3: Covers logical view</td>
</tr>
<tr>
<td>Activity diagram</td>
<td>Compute a quantitative measure of complexity</td>
<td>Activities and connections</td>
<td>Dynamic</td>
<td>No subjective assessment</td>
<td>1: Very simple calculations 2: Considers testing effort in complexity measurement. 3: Covers Use case and process view.</td>
</tr>
</tbody>
</table>

From the summary table it can be concluded that complexity measure using the activity diagram can give a better estimate of complexity as it considers the testing effort by estimating the linearly independent paths from the activity flow graph constructed from the Activity diagram. The independent paths can be used for black box testing to test the paths involving the activities in the path. Test cases need to be designed to test the flow that covers the activities in each path.

IV. METHODOLOGY

The proposed approach intends to calculate the coding complexity aspect from the activity diagram representation of the system/class/function without going into the coding aspect contrary to other complexity measures proposed by McCabe, Halstead or Zage. This complexity measure can be considered as a significant indicator for analysing and implementing the architectural aspects related to quality. This measure can be an indicator for estimating the time and effort to build the system. Activity diagrams are used to represent the dynamic aspect of the system. Study on the mapping of UML diagrams with the 4 + 1 view of architecture show that there are different possible mappings proposed in different works [29]. Mapping proposed in the whitepaper “Applying 4+1 View Architecture with UML 2” [30] is considered for this work where activity diagrams are shown to represent the process view of the system. Activity diagrams are constructed to represent the activities either of a class or the system as a whole. In cases if the activity diagram represents the activities of a class, the system complexity is calculated by the aggregation of the individual class complexities of the system. In cases where the activity diagram represents the activities of the system, the complexity value calculated represents the system complexity. The proposed approach starts with the activity diagram, transforms the diagram to a flow graph and then assesses the complexity. To draw the flow graph, termed as Activity Flow Graph (AFG) the two elements – activities and conditions from the activity diagram are considered to be represented in the form of a graph. The graphical representation of the activity diagram is constructed with nodes and connections between the nodes. Every activity and decision point in the activity diagram is represented as a node in the graph. The start and the termination point of the activity diagram are also represented as nodes in the graph. The connection between the activities and the decision points in the diagram is activity diagram is represented by connecting the respective nodes in the graph. Every connection is considered as an edge in the diagram. A node is represented as a circle and edges are represented as directed arrows. Once the AFG is constructed from the activity diagram, the two metrics required to estimate the complexity are determined. The following two metrics are considered:

- Number of nodes and
- Number of linearly independent paths.
Complexity Assessment based on UML-Activity Diagram

The number of nodes metrics is chosen as it represents the number of activities and conditions that are directly related to coding complexity in terms of Lines of Code (LOC). More activities and conditions represents increased functionality and decision making points which require more effort in coding and leads to increase in LOC. This increase in LOC further gives rise to greater possibility of errors, defects and faults. Moreover, testing effort will also increase due to more LOC in connection to white-box testing. The second metrics that is the number of linearly independent paths introduces complexity from the testing point of view as more test cases needs to be generated to test each path. Here the testing is related to the black box testing aspect as each path will point to a set of actions/task and decision to reach the termination point.

The proposed complexity measure is defined as:

$$C = N + IP \quad \text{(1)}$$

Where,
- $C$ is the complexity measure,
- $N$ is the no. of nodes and
- $IP$ is the no. of linearly independent paths.

As defined earlier in software engineering there are three ways to calculate the number of linearly independent paths.

i) Number of edges – Number of nodes + 2
ii) Number of predicate nodes + 1
iii) Number of regions

For our purpose we have considered the method that uses nodes and edges.

$$IP = E - N + 2 \quad \text{(2)}$$

Where,
- $E$ is the no. of edges and
- $N$ is the number of nodes

A Case study is given here to illustrate the conversion of activity diagram into an AFG and then calculate the complexity measure $C$ from the metrics values collected from the graph.

**CASE STUDY:** Activity diagram of an order management system converted to corresponding Activity Flow Graph and value of $C$ calculated.

For the given case:

AGF notations for the activity diagram are:
- Node S represents the start of the activity.
- Nodes A1 to A4 represents the four activities.
- Nodes D1 and D2 represents the two decision points.
- Node T represents the termination of the activity.

From the AFG, it is found:

- Number of nodes, $N = 6$ (Nodes S and T are not considered as they do not introduce any complexity to the system)
- Number of Edges, $E = 9$
- $IP = 9 - 6 + 2 = 5$
- $C = 6 + 5 = 11$

In this case $C=11$, which reflects the complexity index of software development, calculated from the activity diagram. Every view represents one aspect of the system, and this quantitative measure calculated aims to quantify the amount of complexity in software development from coding as well as testing point of view. Higher value indicates more complexity due to increase in size which results in more effort in cognitive process, coding as well as testing.

**V. IMPLEMENTATION OF THE APPROACH ON A REAL-LIFE TOOL**

The proposed approach is used to calculate the complexity of a real-life tool called AppBuilder developed at National Informatics Centre*, which basically tries to generate various application objects (e.g. Forms, Menus etc.) from database metadata. There are eight activity diagrams representing the system. The activity diagram for generating struts based form is given here and its corresponding AFG is generated to calculate the complexity measure.
The complexity measure for the struts based form is calculated using Eq (1) defined earlier.

No. of nodes (N) = 10
No. of linearly independent paths (IP) = 13 – 10 + 2 = 5

C = 10 + 5 = 15

The C value for all the other activity diagrams are estimated using the proposed approach. The values are given below:

For Javabean class:
No. of nodes = 6.
No. of Edges = 8.

For EntityManager class:
No. of nodes = 11.
No. of Edges = 13.
IP = 13 – 11 + 2 = 4.

Therefore A4 = 11 + 4 = 15.

For Action class:
No. of nodes = 8.
No. of Edges = 11.
IP = 11 – 8 + 2 = 5.

Therefore A5 = 8 + 5 = 13.

For configproperties.properties file:
No. of nodes = 7.
No. of Edges = 9.
IP = 9 – 7 + 2 = 4.

Therefore A6 = 7 + 4 = 11.

For global.properties file:
No. of nodes = 7.
No. of Edges = 9.
IP = 9 – 7 + 2 = 4.

Therefore A7 = 7 + 4 = 11.

For struts.xml file:
No. of nodes = 6.
No. of Edges = 8.
IP = 8 – 6 + 2 = 4.

Therefore A8 = 6 + 4 = 10.

For web.xml file:
No. of nodes = 5.
No. of Edges = 7.
IP = 7 – 5 + 2 = 4.

Therefore A9 = 5 + 4 = 9.

The following table represents the C value for the different activity diagrams used in the project.

<table>
<thead>
<tr>
<th>Entity name</th>
<th>C-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Javabeanclass</td>
<td>10</td>
</tr>
<tr>
<td>EntityManagerclass</td>
<td>13</td>
</tr>
<tr>
<td>Actionclass</td>
<td>15</td>
</tr>
<tr>
<td>Configpropertiesfile</td>
<td>11</td>
</tr>
<tr>
<td>Globalpropertiesfile</td>
<td>11</td>
</tr>
<tr>
<td>Strutsform</td>
<td>15</td>
</tr>
<tr>
<td>Struts</td>
<td>10</td>
</tr>
<tr>
<td>Web</td>
<td>9</td>
</tr>
</tbody>
</table>

From table 2 it is observed that the C-value varies from 9 to 15 which signify that there is difference in complexity in case of the listed entities of the system. The creator of the system commenting on the observation concludes that the identified classes having highest C-value of 15 and 13 were indeed the most complex entities to develop. In addition he remarks that, as the project basically tries to generate code for an application in the Model-View-Controller (MVC) architecture, the identified three entities take care of the View (Strutsform), Model (EntityManager) and Controller (Actionclass).
VI. ANALYSIS AND RESULTS

In order to ascertain and obtain a quantitative measure of extent of complexity, further experimentation on the source code of the above mentioned entities was conducted using the SourceMonitor tool. The aim of this experimentation was to explore if there could be a quantitative measure that relates the calculated C-value in terms of existing complexity metrics (LOC). The SourceMonitor tool, version 3.5.3.328 by

Campwood is used for this purpose. SourceMonitor is a tool for coders intended for use by programmers as they write code, to expose to them information about the code they are creating. As the project has html, java and xml code, this freeware tool was suitable to estimate the complexity and other parameters like LOC and methods. The table below summarises the different parameter values collected from the analysis.

### Table 3: The values of the parameters of the entities

<table>
<thead>
<tr>
<th>Entity name</th>
<th>Estimated C value (from activity diagrams)</th>
<th>LOC (from tool)</th>
<th>Methods (from tool)</th>
<th>Complexity (from tool)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JavaBeanClass(java)</td>
<td>10</td>
<td>56</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>EntityManagerClass(java)</td>
<td>13</td>
<td>137</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>ActionClass(java)</td>
<td>15</td>
<td>204</td>
<td>14</td>
<td>8</td>
</tr>
<tr>
<td>Configfile(java)</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Globalfile(java)</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strutsform(html)</td>
<td>15</td>
<td>104</td>
<td>99 (tags)</td>
<td>2.9 (complex tags)</td>
</tr>
<tr>
<td>Struts(xml)</td>
<td>10</td>
<td>14</td>
<td>8 (tags)</td>
<td>0</td>
</tr>
<tr>
<td>Web.xml</td>
<td>9</td>
<td>19</td>
<td>11 (tags)</td>
<td>0</td>
</tr>
</tbody>
</table>

As the coding is in different languages, the groups are shown in the table using different colours. Analysis is done based on the two groups - java coded entities and html/xml coded entities. As no parameters are evaluated by the tool for two entities, they are not considered for inference purpose in this study. From the values shown in table 2, it is observed that there is no pattern followed in case of the java class C values and the complexity calculated using the tool. The C values and the LOC values show an increasing trend. Increase in C value shows an increase in LOC. The second part of the analysis is to find a relationship between the C-value and the LOC obtained from the source code using the tool. The table below summarises the analysis to establish a relationship between C and LOC.

### Table 4: Relationship of C and LOC

<table>
<thead>
<tr>
<th>Difference in C-value (X)</th>
<th>Difference in LOC(Y)</th>
<th>Y/X</th>
<th>Relationship of C with LOC(Z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>13-10=3</td>
<td>137-56=8</td>
<td>81/3</td>
<td>Z=(Σ Y/X )/3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>= (27+34+30)/3=30.33≈30.</td>
</tr>
<tr>
<td>15-13=2</td>
<td>204-137=67</td>
<td>67/2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>≈ 34</td>
</tr>
<tr>
<td>15-10=1</td>
<td>204-56=148</td>
<td>148/5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>≈ 29.6 ≈ 30</td>
</tr>
</tbody>
</table>

From Table 4 it can be observed that an increase of one in C value could approximately increase the LOC by 30. It can be inferred that C can be used as an indicator of the size of the system or function or class in case of java code. Other observations in context of html/xml code do not show any significant trend except that a higher value of C implies more number of tags and LOC used for development.

The results concluded after implementing the approach on the activity diagrams of a real life tool and then analysing the source code of the entities can be summarised as follows:

- There is a direct relationship between the C-value calculated using the proposed method with the complexity in developing the system/class/function which the activity diagram represents. Higher C value indicates higher complexity.
- From the observations it is found that, the C-value calculated from the activity diagram can be used to estimate the approximate LOC of the system, class or function for which the activity diagram is drawn.
- Increase in C-value (above 10) by one is equivalent to increase in 30 lines of LOC.

VII. CONCLUSION AND FUTURE SCOPE

The present work on complexity measure proposes a new method for evaluating the process view of the system considering architectural view mapping with UML diagrams. The representation is in the form of an activity diagram representing the process view which is converted to a flow graph. A complexity value is estimated from the representation using two metrics. As one of the metrics measures the independent paths, the complexity measure considers the testing effort into consideration while estimating the complexity value. As an activity diagram can be drawn for a system, class or function, the same approach can be used to estimate the complexity of a system/class/function. This complexity value can be used to estimate the time and effort required to develop the system. The measure of complexity can be used in addition to other attributes like modifiability, operability.
maintainability and cost to evaluate the architecture of a system. Implementing the proposed approach on a real life tool, a relationship between the calculated complexity measure and the LOC metrics could be established. This relationship may be used to estimate the size of the system. As size is the basic estimate based on which other project parameters like time, cost and other resources can be estimated, this relationship is significant. The conversion of the activity diagram to AFG provides a basis on which the complexity measure is estimated. This estimation can be further refined to provide more accurate estimates by considering cognitive weights to the nodes and edges. These cognitive weights can consider the subjective aspects in estimation. Further the structural view complexity aspects can be incorporated in refining the complexity measure which is a direction for future work.

REFERENCES