Abstract—The requirement to improve software productivity has promoted the research on software metric technology. There are metrics for identifying the quality of reusable components but the function that makes use of these metrics to find reusability of software components is still not clear. These metrics if identified in the design phase or even in the coding phase can help us to reduce the rework by improving quality of reuse of the component and hence improve the productivity due to probabilistic increase in the reuse level. CK metric suit is most widely used metrics for the object-oriented (OO) software; we critically analyzed the CK metrics, tried to remove the inconsistencies and devised the framework of metrics to obtain the structural analysis of OO-based software components. Neural network can learn new relationships with new input data and can be used to refine fuzzy rules to create fuzzy adaptive system. Hence, Neuro-fuzzy inference engine can be used to evaluate the reusability of OO-based component using its structural attributes as inputs. In this paper, an algorithm has been proposed in which the inputs can be given to Neuro-fuzzy system in form of tuned WMC, DIT, NOC, CBO, LCOM values of the OO software component and output can be obtained in terms of reusability. The developed reusability model has produced high precision results as expected by the human experts.

Keywords—CK-Metric, ID3, Neuro-fuzzy, Reusability.

I. INTRODUCTION

The aim of Object Oriented (OO) Metrics is to predict the quality of the object oriented software products. Various attributes, which determine the quality of the software, include maintainability, defect density, fault proneness, normalized rework, understandability, reusability etc. The requirement nowadays is to explore the relation of the reusability attributes with the metrics and to find how these metrics collectively determine the reusability of the software component. To achieve both the quality and productivity objectives, it is always recommended to go for the software reuse that not only saves the time taken to develop the product from scratch but also delivers the almost error free code, as the code is already tested many times during its earlier reuse.

A great deal of research over the past several years has been devoted to the development of methodologies to create reusable software components and component libraries, where there is an additional cost involved to create a reusable component from scratch. That additional cost could be avoided by identifying and extracting reusable components from the already developed large inventory of existing systems. But the issue of how to identify good reusable components from existing systems has remained relatively unexplored. Our approach, for identification and evaluation of reusable software, is based on software models and metrics. As the exact relationship between the attributes of the reusability is difficult to establish, a Neural Network approach could serve as an economical, automatic tool to generate reusability ranking of software [1] [2] by formulating the relationship based on its training. When one designs with Neural Networks alone, the network is a black box that needs to be defined; this is a highly compute-intensive process. One must develop a good sense, after extensive experimentation and practice, of the complexity of the network and the learning algorithm to be used. Neural nets and fuzzy systems, although very different, have close relationship: they work with impression in a space that is not defined by crisp, deterministic boundaries [3]. Neural network can be used to define fuzzy rules for the fuzzy inference system. A neural network is good at discovering relationships and pattern in the data, so neural network can be used to preprocess data in the fuzzy system. Furthermore, neural network that can learn new relationships with new input data can be used to refine fuzzy rules to create fuzzy adaptive system. With the objective of taking advantage of the features of the both [4], we used Neuro-Fuzzy approach to economically determine reusability of OO-based software components in existing systems as well as the reusable components that are in the design phase. Inputs to Neuro-fuzzy system, are provided in form of tuned WMC, DIT, NOC, CBO, LCOM values of the OO-based software component and output is be obtained in terms of reusability.

II. BACKGROUND

It was Selby [5], who tried to identify a number of characteristics of those components, from existing systems, that are been reused at NASA laboratory and reported that the developers there has achieved a 32 percent reusability index. Dunn and Knight in 1991[6] also experimented and reported the usefulness of reusable code scavenging. Chen, Nishimoto and Ramamooorthy briefly discuss the idea of subsystem extraction by using code information stored in a relational
A. OO-Based Metric Framework

As CK metric suit is able to target all the essential attributes of OO-based software, as mentioned by Selby [13] in his latest findings, so we analyzed, refined and used following metrics of CK metric suit to explore different structural dimensions of a class.

1) Weighted methods per class (WMC): According to this metric if a Class C, has n methods and \( c_1, c_2 \ldots c_n \) be the complexity of the methods, then \( \text{WMC}(C) = c_1 + c_2 + \ldots + c_n \).

McCabe’s complexity metric is chosen for calculating the complexity values of the methods of a class. The value is normalized so that nominal complexity for a method takes on a value of 1.0. If all method complexities are considered to be unity, then \( \text{WMC} = n \) i.e. the number of methods existing in that class [14][15].

We have used “tuned WMC” (TWMC) measure as input to the neuro-fuzzy inference engine by restricting the WMC value in between 0 and 1 with help of sigmoidal function as shown in (1).

\[
f(x,a,c) = \frac{1}{1 + e^{-a(x-c)}}
\]

Where \( a=10 \) and \( c=0.5 \).

2) Depth of inheritance tree (DIT): According to this metric, Depth of inheritance of a class is “the maximum length from the node to the root of the tree”. More is the depth of the inheritance tree greater the reusability of the class corresponding to the root of that tree as the class properties are shared by more derived classes under that class. So there too much depth dilutes the abstraction. So there is a need to set the minimum & maximum DIT value for a class as an contribution towards the reusability [14][15].

The definition of DIT is ambiguous when multiple inheritance and multiple roots are present as the alternative length of the path is not being considered in case of multiple inheritance. If we add all the ancestor classes coming in common path to the ancestor classes coming in alternative paths then that will be the true representation of the theoretical basis of the DIT metric [16].

We have used “lack of tuned degree of inheritance” (LTDIT) measure as input to the Neuro-fuzzy inference system, in order to restrict the input value between 0 and 1.

3) Number of Children (NOC): According to this metric Number of children (NOC) of a class is the number of immediate sub-classes subordinated to a class in the class hierarchy. So greater is the value of NOC greater will be the reusability of the parent class. Hence there should be some minimum value of NOC for a parent class for its reusability [14][15].

Theoretical basis of NOC metric relates to the notion of scope of properties. It is a measure of how many sub-classes are going to inherit the methods of the parent class [14].
definition of NOC metric gives the distorted view of the system as it counts only the immediate sub-classes instead of all the descendants of the class. So according to [16] the NOC value of a class, say class i, should reflect all the subclasses that share the properties of that class as shown in (2).

\[ NOC(i) = N + \sum \text{Subclasses} \times NOC(i) \tag{2} \]

Where N is the total number of immediate subclasses of class i. In order to restrict the input value between 0 and 1, we have used “lack of tuned Number of Children” (LTNOC) measure as input to the Neuro-fuzzy inference system.

4) Coupling Between Object Classes (CBO): According to this metric “Coupling Between Object Classes” (CBO) for a class is a count of the number of other classes to which it is coupled. Theoretical basis of CBO relates to the notion that an object is coupled to another object if one of them acts on the other, i.e. methods of one object use methods or instance variables of another. Here we are restricting the unidirectional use of methods or instance variables of another object by the object of the class whose reusability is to be measured. As Coupling between Object classes increases, reusability decreases and it becomes harder to modify and test the software system. So there is the need to set some maximum value of coupling level for its reusability. If the value of CBO for a class is beyond that maximum value then the class is said to be non-reusable[14][15].

In order to restrict the input value between 0 and 1, we have used “lack of CBO” (LCBO) measure as input to the Neuro-fuzzy inference system.

5) Lack of Cohesion in Methods (LCOM): Consider a Class C1 with n methods M1 , M2 , ..., Mn . Let \{Ij\} be set of instance variables used by method Mj . There are n such sets \{I1\}, \{I2\}, ..., \{In\}. Let P = \{(I, Ij) | I \& Ij \in \emptyset\} and Q = \{(I, Ij) | I \& Ij \notin \emptyset\}. If all n sets \{I1\}, \{I2\}, ..., \{In\}, are \emptyset then let P = \emptyset [4]. Lack of Cohesion in Methods (LCOM) of a class can be defined as

\[ LCOM = |P| - |Q|, \text{if } |P| > |Q| \]

LCOM = 0 otherwise

The high value of LCOM indicates that the methods in the class are not really related to each other and vice versa means less reusability otherwise low value of LCOM depicts high internal strength of the class which results into high reusability. So there should be some maximum value of LCOM after which class becomes non-reusable [14][15].

We have used “tuned LCOM” (TLCOM) measure as input to the neuro-fuzzy inference engine by restricting the LCOM value in between 0 and 1 with help of sigmoidal function as shown in (3).

\[ f(x,a,c) = \frac{1}{1 + e^{-\frac{a(x-c)}{w}}} \tag{3} \]

Where \(a=4\) and \(c=1.5\).

B. Neuro-Fuzzy System’s Architecture

The fuzzy logic approach is beneficial for measuring the reusability of a software component as the conventional model based approaches are difficult to be implemented. Unfortunately, with the increase in the complexity of the problem being modeled and unavailability of the precise relationship among various constituents for measuring the reusability, has led to rely on another approach which is mostly known as neuro-fuzzy or fuzzy-neuro approach. It has the benefits of both neural networks and fuzzy logic. The neuro-fuzzy hybrid system combines the advantages of fuzzy logic system, which deal with explicit knowledge that can be explained and understood, and neural networks, which deal with implicit knowledge, which can be acquired by learning.

A fuzzy system can be considered to be a parameterized nonlinear map, called \(f\), which can be expressed as in (4).

\[ f(x) = \frac{\sum \mu_{Ai}}{\sum \mu_{Bi}} \tag{4} \]

Where \(y^l\) is a place of output singleton if Mamdani reasoning is applied or a constant if Sugeno reasoning is applied. The membership function \(\mu_{Ai}(x_i)\) corresponds to the input \(x_i=[x_1, x_2, x_3, ..., x_n]\) of the rule \(l\). The “and” connective in the premise is carried out by a product and defuzzification by the center-of-gravity method [4]. Consider a Sugeno type of fuzzy system having the rule base

Rule1: If x is A1 and y is B1, then \(f_1 = p_1x + q_1y + r_1\)
Rule2: If x is A2 and y is B2, then \(f_2 = p_2x + q_2y + r_2\)

Let the membership functions of fuzzy sets \(A_i, B_i, i=1,2,\) be \(\mu_{Ai}, \mu_{Bi}\).

1) Evaluating the rule premises results values as shown in (5).

\[ w_i = \mu_{Ai}(x) \times \mu_{Bi}(y) \tag{5} \]

where \(i = 1,2\) for the rule rules stated above.

2) Evaluating the implication and the rule consequences gives values as shown in (6)-(9).

\[ f(x,y) = \frac{w_1 f_1(x,y) + w_2 f_2(x,y)}{w_1 + w_2} \tag{6} \]

Or

\[ f = \frac{w_1f_1 + w_2f_2}{w_1 + w_2} \tag{7} \]

Let

\[ \frac{w_1}{w_2} = \frac{w_1}{w_1 + w_2} \tag{8} \]

then \(f\) can be written as
These all computations can be presented in a diagram form [4] as shown in the fig. 1(a) and 1(b).

\[ f = w_1 f_1 + w_2 f_2 \]  
(9)

In the neuro-fuzzy inference system using a given input/output data set, we have constructed a fuzzy inference system (FIS) whose membership function parameters are tuned (adjusted) using a hybrid method consisting of back-propagation form of the steepest descent method for the parameters associated with the input membership functions, and least squares estimation for the parameters associated with the output membership functions. As a result, the training error decreases, at least locally, throughout the learning process. Therefore, the more the initial membership functions resemble the optimal ones, the easier it will be for the model parameter training to converge.

In the neuro-fuzzy inference system using a given input/output data set, we have constructed a fuzzy inference system (FIS) whose membership function parameters are tuned (adjusted) using a hybrid method consisting of back-propagation form of the steepest descent method for the parameters associated with the input membership functions, and least squares estimation for the parameters associated with the output membership functions. As a result, the training error decreases, at least locally, throughout the learning process. Therefore, the more the initial membership functions resemble the optimal ones, the easier it will be for the model parameter training to converge.

IV. IMPLEMENTATION

In the fuzzy Inference system Linguistic variables are then assigned to the input parameters based on their values. The assignment of the linguistic variables depends on the range of the input measurement.

Values to the linguistic variables of TWMC are assigned in terms of complexity of the software component. TWMC is assigned three linguistic variables LOW, MEDIUM and HIGH in the range of 0 to 1.

Values to the linguistic variables of LTDIT (Lack of tuned degree of Inheritance) are assigned in terms of level of inheritance for the software component under consideration. LTDIT is assigned three linguistic variables LOW, HIGH and MEDIUM in the range of 0 to 1.

Values to the linguistic variables of LTNOC (Lack of tuned Number of Children) are assigned in terms of number of children of the software component in the inheritance tree. Quality attribute LTNOC is assigned three linguistic variables LOW, MEDIUM and HIGH in the range of 0 to 1.

Values to the linguistic variables of LCBO (Lack of Coupling between Objects) are assigned in terms of coupling between objects of the software. LCBO is assigned three linguistic variables LOW, MEDIUM and HIGH in the range of 0 to 1.

Values to the linguistic variables of TLCOM (tuned Lack of Cohesion) are assigned in terms of Cohesion in methods of the class of object-oriented software TLCOM is assigned three linguistic variables LOW, MEDIUM and HIGH in the range of 0 to 1.

Values to the linguistic variables of Reusability are assigned in terms of “how reusable the software component is?” The output membership functions are only linear or constant for Sugeno-type fuzzy inference. Reusability is assigned six linguistic variables PERFECT, HIGH, MEDIUM, LOW, VERY-LOW and NIL as constants in the range of 0-1.

A network-type structure similar to that of a neural network, which maps inputs through input membership functions and associated parameters, and then through output membership functions and associated parameters to outputs, can be used to interpret the input/output map is shown in the fig. 3. The parameters associated with the membership functions will change through the learning process. The computation of these parameters (or their adjustment) is facilitated by a gradient vector, which provides a measure of how well the fuzzy inference system is modeling the input/output data for a given set of parameters. Once the gradient vector is obtained, any of several optimization routines could be applied in order to adjust the parameters so as to reduce some error measure. The Error Tolerance is used to create a training stopping criterion, which is related to the error size. The training will stop after the training data error remains within this tolerance. This is set to 0 as we don’t know how training error is going to behave.
The Training of the neuro-fuzzy system is performed using Training Data, as mentioned in appendix-I, for 1000 iterations and the training error reduces after each iteration as shown by the fig. 4 and stabilities at the error value of 0.041969, so at this point the network is said to be converged.

V. CONCLUSION

During the testing phase, when the developed system is tested against the testing data, as mentioned in appendix-II, Average Testing error 0.039364 is obtained. The plot between the actual output and the expected output is shown in fig. 5. As the actual output produced by the “Reusability Evaluation Model” is close to the expected output, so the system can be recommended for Automatic identification potential reusable object oriented software components from the legacy systems and evaluating the quality of developed or developing reusable components for better productivity and quality.
TABLE II
TESTING DATA FOR NEURO-FUZZY SYSTEM

<table>
<thead>
<tr>
<th>TWMC</th>
<th>LTDT</th>
<th>LTNO</th>
<th>LCBO</th>
<th>TCLON</th>
<th>REUSABILITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.83</td>
<td>0.97</td>
<td>0.79</td>
<td>0.63</td>
<td>0.21</td>
<td>0.07</td>
</tr>
<tr>
<td>0.32</td>
<td>0.29</td>
<td>0.18</td>
<td>0.78</td>
<td>0.13</td>
<td>0.87</td>
</tr>
<tr>
<td>0.13</td>
<td>0.80</td>
<td>0.87</td>
<td>0.43</td>
<td>0.84</td>
<td>0.11</td>
</tr>
<tr>
<td>0.85</td>
<td>0.79</td>
<td>0.92</td>
<td>0.22</td>
<td>0.14</td>
<td>0.05</td>
</tr>
<tr>
<td>0.45</td>
<td>0.20</td>
<td>0.19</td>
<td>0.75</td>
<td>0.12</td>
<td>0.80</td>
</tr>
<tr>
<td>0.40</td>
<td>0.56</td>
<td>0.44</td>
<td>0.68</td>
<td>0.23</td>
<td>0.62</td>
</tr>
<tr>
<td>0.62</td>
<td>0.23</td>
<td>0.13</td>
<td>0.93</td>
<td>0.24</td>
<td>0.93</td>
</tr>
<tr>
<td>0.66</td>
<td>0.45</td>
<td>0.67</td>
<td>0.64</td>
<td>0.20</td>
<td>0.68</td>
</tr>
<tr>
<td>0.91</td>
<td>0.88</td>
<td>0.88</td>
<td>0.13</td>
<td>0.20</td>
<td>0.08</td>
</tr>
<tr>
<td>0.69</td>
<td>0.64</td>
<td>0.59</td>
<td>0.93</td>
<td>0.19</td>
<td>0.62</td>
</tr>
<tr>
<td>0.0</td>
<td>0.83</td>
<td>0.89</td>
<td>0.39</td>
<td>0.93</td>
<td>0.14</td>
</tr>
<tr>
<td>0.71</td>
<td>0.80</td>
<td>0.77</td>
<td>0.19</td>
<td>0.14</td>
<td>0.02</td>
</tr>
<tr>
<td>0.41</td>
<td>0.42</td>
<td>0.40</td>
<td>0.46</td>
<td>0.71</td>
<td>0.57</td>
</tr>
<tr>
<td>0.87</td>
<td>0.76</td>
<td>0.89</td>
<td>0.21</td>
<td>0.23</td>
<td>0.04</td>
</tr>
</tbody>
</table>

ACKNOWLEDGMENT
The authors like to express their gratitude towards Dr. S. B. Singh (Principal, G.N.D.E.C., Ludhiana) and Dr. H. K. Grewal, HOD (CSE & IT), G.N.D.E.C, Ludhiana for provision of laboratory facilities.

REFERENCES


