An Intelligent Approach to Software Cost Prediction

Xishi Huang, Danny Ho\textsuperscript{1}, Luiz F. Capretz, Jing Ren

Dept. of ECE, University of Western Ontario, London, Ontario, N6G 1H1, Canada
1Toronto Design Center, Motorola Canada Ltd., Markham, Ontario, L6G 1B3, Canada

{xhuang22, lcapretz, jren2}@uwo.ca, Danny.Ho@motorola.com

October 22, 2003
Agenda

- Problem
- The Novel Neuro-Fuzzy Model
- Validation by Industry Project Data
- Conclusion and Future Work
Software Cost Estimation

Motivation:
- Many software projects go over time and budget
- Software development has become an essential investment for many organizations

Problem Statement:
- We use project information such as software size and other attributes to predict software cost or effort

Characteristics of Software Cost Estimation:
- Complex nonlinear relationships between software development cost and cost drivers
- Imprecise and uncertain measurement
- Rapid change of software technology and processes
Agenda

- Problem
- A Novel Neuro-Fuzzy Model
- Validation by Industry Project Data
- Conclusion and Future Work
A Novel Neuro-Fuzzy COCOMO Model

COCOMOII MODEL

\[ A \times (\text{Size})^{B + 0.01 \sum_{i=1}^{5} S_{Fi}} \times \prod_{i=1}^{17} E_{Mi} \]
COCOMO Model

- COCOMO II Post Architecture Model

\[
Effort = A \times (Size)^{B + 0.01 \times \sum_{i=1}^{5} SF_i} \times \prod_{i=1}^{17} EM_i
\]

where:
- effort is in staff-month (SM)
- A, B are constants
- Size is in KSLOC
- \(SF_i\)'s are scale factors
- \(EM_i\)'s are effort multipliers
Some Notations

<table>
<thead>
<tr>
<th>Fuzzy Sets</th>
<th>Rating Levels</th>
<th>Parameter Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_{i1}$</td>
<td>1 Very Low (VL)</td>
<td>$CD_{i1}$</td>
</tr>
<tr>
<td>$A_{i2}$</td>
<td>2 Low (L)</td>
<td>$CD_{i2}$</td>
</tr>
<tr>
<td>$A_{i3}$</td>
<td>3 Nominal (N)</td>
<td>$CD_{i3}$</td>
</tr>
<tr>
<td>$A_{i4}$</td>
<td>4 High (H)</td>
<td>$CD_{i4}$</td>
</tr>
<tr>
<td>$A_{i5}$</td>
<td>5 Very High (VH)</td>
<td>$CD_{i5}$</td>
</tr>
<tr>
<td>$A_{i6}$</td>
<td>6 Extra High (XH)</td>
<td>$CD_{i6}$</td>
</tr>
</tbody>
</table>
Sub-Model $NF_i$

$CDR_i$: continuous rating value

$CD_i$: corresponding numerical value
Reasoning Phase: Fuzzy rules

- If $CDR_i$ is $A_{i1}(VL)$, then $CD_i = CD_{i1}$,
- If $CDR_i$ is $A_{i2}(L)$, then $CD_i = CD_{i2}$,
- If $CDR_i$ is $A_{i3}(N)$, then $CD_i = CD_{i3}$,
- If $CDR_i$ is $A_{i4}(H)$, then $CD_i = CD_{i4}$,
- If $CDR_i$ is $A_{i5}(VH)$, then $CD_i = CD_{i5}$,
- If $CDR_i$ is $A_{i6}(XH)$, then $CD_i = CD_{i6}$.

$(i,k), \ i=1,2,\ldots,22, \ k=1,2,\ldots,6$
Sub-Model $NF_i$ (Cont’d)

- **Layer 1:**
  \[ O^1_k = \mu_{ik} (x) \]

- **Layer 2:**
  \[ w_k = \prod \mu_{ikj} (x) \]

- **Layer 3:**
  \[ \overline{w}_k = \frac{w_k}{\sum w_j} \]

- **Layer 4:**
  \[ O^4_k = \overline{w}_k CD_{ik} \]

- **Layer 5:**
  \[ CD_i = \sum_k \overline{w}_k CD_{ik} \]
Interpretation of sub-model $NF_i$

Linear interpolation:

$\mu_{ik}(x) = \begin{cases} 
  x - (k - 1), & k - 1 \leq x \leq k \\
  (k + 1) - x, & k \leq x \leq k + 1 \\
  0, & \text{otherwise}
\end{cases}$

$$CD_i = CD_{ik} + (CD_{ik+1} - CD_{ik})(CDR_i - k)$$

$k \leq CDR_i \leq k + 1, k = 1, 2, \ldots, 6$
Learning Algorithms

Optimization problem:

\[ E = \sum_{n=1}^{NN} \frac{1}{2} w_n \left( \frac{E_n - E_{dn}}{E_{dn}} \right)^2 \]

subject to monotonic constraints:

\[ CD_{i_1} \leq CD_{i_2} \leq CD_{i_3} \leq CD_{i_4} \leq CD_{i_5} \leq CD_{i_6}, i \in I_{INC} \(CD\) \]

\[ CD_{i_1} \geq CD_{i_2} \geq CD_{i_3} \geq CD_{i_4} \geq CD_{i_5} \geq CD_{i_6}, i \in I_{DEC} \(CD\) \]

Learning Algorithm:

\[ CD_{\frac{l+1}{ik}} = CD_{\frac{l}{ik}} - \alpha \frac{\partial E}{\partial \ CD_{\frac{l}{ik}}} \]
Advantage of Continuous Rating Values

Two similar projects P1 and P2 have the same nominal effort, say 100 staff-months.

The COCOMO II.2000 model,

- For P1, 203 staff-months
- For P2, 2886 staff-months!

The difference is over 14 times.

Our Model: 809 staff-months for both projects
Agenda

- Problem
- A Novel Neuro-Fuzzy Model
- Validation by Industry Project Data
- Conclusion and Future Work
Validation by Industry Project Data

Sources of Project Data

- Industrial project data: 6 projects
- COCOMO81 database: 63 projects
Validation Results

- Case I. Learning with all project data
- Case II. Learning with part of project data
- Case III. Use larger weights for local data
- Case IV. Learning without monotonic constraints
### Effort estimation for all 69 project data points

<table>
<thead>
<tr>
<th>ARE</th>
<th>COCOMO81 Model</th>
<th>Neuro-Fuzzy Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case I</td>
<td>Case II</td>
<td>Case III</td>
</tr>
<tr>
<td>PERC</td>
<td>IMPRV</td>
<td>PERC</td>
</tr>
<tr>
<td>20%</td>
<td>71%</td>
<td>86%</td>
</tr>
<tr>
<td>30%</td>
<td>81%</td>
<td>92%</td>
</tr>
</tbody>
</table>
**Effort estimation for industrial project data**

<table>
<thead>
<tr>
<th>Project No</th>
<th>Actual Effort</th>
<th>COCOMO81 Model</th>
<th>Neuro-Fuzzy Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Estimate</td>
<td>Error</td>
</tr>
<tr>
<td>P1</td>
<td>638.0</td>
<td>745.0</td>
<td>16%</td>
</tr>
<tr>
<td>P2</td>
<td>185.0</td>
<td>167.3</td>
<td>-9%</td>
</tr>
<tr>
<td>P3</td>
<td>332.0</td>
<td>322.0</td>
<td>-3%</td>
</tr>
<tr>
<td>P4</td>
<td>619.9</td>
<td>651.5</td>
<td>5%</td>
</tr>
<tr>
<td>P5</td>
<td>64.8</td>
<td>63.4</td>
<td>-2%</td>
</tr>
<tr>
<td>P6</td>
<td>76.6</td>
<td>72.2</td>
<td>-5%</td>
</tr>
</tbody>
</table>
Agenda

- Problem
- A Novel Neuro-Fuzzy Model
- Validation by Industry Project Data
- Conclusion and Future Work
Conclusion

- Propose a novel neuro-fuzzy COCOMO model
  - Neuro-fuzzy COCOMO structure
  - Monotonic constraints
  - Learning algorithm
  - Fuzzy rules
Conclusion (Cont’d)

Distinguishing Features of the Proposed Model:

- Learning ability
- Robust to imprecise and uncertain inputs
- Good Interpretability
- Knowledge integration
- Reduced number of learning parameters
- Good generalization
- Local learning
Future Work

Extending the neuro-fuzzy approach to

- Other cost estimation models, e.g. SLIM
- Quality estimation models, e.g. COQUALMO
References

THANKS!
Any Questions?