ENNA: Software Effort Estimation Using Ensemble of Neural Networks with Associative Memory

Yigit Kultur, Burak Truhan, and Ayse Basar Bener

Symposium on Foundations of Software Engineering, 2008

Hyunsik Cho
Mar 9, 2011
Contents

- Introduction
- Overall approach
- Ensemble of Neural Networks with Associative Memory (ENNA)
- Experiment
- Conclusion
- Discussion
Software effort estimation

Accurate effort estimation is an important factor for successful software project management\(^1\)

- Underestimating causes schedule-delay, over-budget, poor software quality, and angry customers
- Overestimation causes a waste of software development resources

\(^1\) F. Heemstra, “software cost estimation,” *Information and software Technology*, pp. 1-14, 1992
Machine learning (ML) methods have been preferred over parametric models

- ML methods are more flexible to calibrate the model
  - Parametric models have pre-specified formula
  - Parametric models can only be calibrated manually
- ML methods allow the learning from previous situations
- Neural Networks method has been widely used
  - It can model a complex set of relationships between the dependable variable and independent variables

- **Parametric models**: COCOMO, COCOMO II, Function Points
- **ML methods**: Case-based reasoning, regression tree, genetic algorithms, neural networks
Neural Networks (NN) for effort estimation

- Commonly used type
  - Feed-forward three-layer Perceptron with Backpropagation learning algorithm and Sigmoid activation function
Motivation

- Dataset in software estimation domain is smaller than dataset in other domains
  - NN are hard to make accurate estimations with small dataset

- NN are unstable structures
  - Small changes in the training set can cause large difference
    - Because learning algorithms has high variance

- NN are memoryless structures
  - Once projects data is trained, NN do not need those any more

- Overtraining of NN has negative impact on accuracy
  - Overtraining leads to poor generalization
Goal of this paper

- Propose ensemble of neural networks with associative memory to increase the estimation performance of NN
  - Use multiple neural networks rather than a single one to alleviate unstable structure
  - Use similar past project explicitly to decrease the bias
Overall approach

Step 1
Historical dataset

New project

Step 2
NN 1
NN 2
NN 20

Effort Ensemble

Step 3
Bias calculation using nearest neighbor

Bias Estimate

Estimated effort
Ensemble of Neural Networks (ENN)

- Training process
  - Use bootstrapping method to alleviate small datasets

* Number of projects

© KAIST SE LAB 2011
ENN (Cont’)

- Test process

- Effort made by each NN may deviate much from each other
  - Some NN would have fallen into local minimum
  - Some NN are inaccurately trained by random chosen training set

Detect largest group of nearby results and use mean of this group
ENNA (4/6)

- **ENN (Cont’)**
  - **ART clustering algorithm**
    - Start with one single cluster
    - Add new cluster if $|\text{input value} - \text{center}_i| > \text{vigilance}$ for all $i$
    - Included in cluster$_i$ if $|\text{input value} - \text{center}_i| < \text{vigilance}$

  - **Example**

  - **Largest group:** $C_3$
  - $\text{Effort}_{\text{Ensemble}} = \sum \frac{\text{Effort}_i}{N_c}, \forall \text{Effort}_i \in C$  \(\text{(C: largest group)}\)
ENN with Associative memory

- Use Associative memory to correct bias of NN model
  - Estimate bias of new project using bias of similar past projects

\[
Bias_{\text{estimated}} = \frac{1}{k} \sum_{i \in N_k} \text{Effort}_{\text{actual}}(i) - \text{Effort}_{\text{ensemble}}(i)
\]
Final estimation

\[ \text{Effort}_{\text{final}} = \text{Effort}_{\text{ensemble}} + \text{Bias}_{\text{estimated}} \]

- \( \text{Effort}_{\text{ensemble}} \): New project’s estimated effort using ensemble of Neural Networks
- \( \text{Bias}_{\text{estimated}} \): New project’s bias derived from similar past projects
Data set preparation

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Source</th>
<th>Size</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>NASA</td>
<td>NASA projects</td>
<td>60</td>
<td>15</td>
</tr>
<tr>
<td>NASA 93</td>
<td>NASA projects</td>
<td>93</td>
<td>15</td>
</tr>
<tr>
<td>USC</td>
<td>USC projects</td>
<td>63</td>
<td>15</td>
</tr>
<tr>
<td>SDR</td>
<td>Softlab Data Repository at Bogazici University</td>
<td>24</td>
<td>17</td>
</tr>
<tr>
<td>Desharnais</td>
<td>Canadian software house projects</td>
<td>77</td>
<td>11</td>
</tr>
</tbody>
</table>

Accuracy measure’s criteria

\[
MRE(i) = \left| \frac{e_p(i) - \hat{e}_p(i)}{e_p(i)} \right|
\]

- \(e_p(i)\) : Actual effort value of \(i^{th}\) project
- \(\hat{e}_p(i)\) : Estimated effort value of \(i^{th}\) project

\[
MMRE = mean(MRE) = \frac{1}{n} \sum_{i=1}^{n} MRE(i)
\]

\[
MdMRE = median(MRE)
\]

\[
PRED(25\%) = \frac{k}{n}
\]

- \(n\) : Number of projects in historical dataset
- \(k\) : Number of projects whose MRE is less than 25%
Validation design

Goal
- Compare estimation accuracy of proposed models (ENN, ENNA) with other estimation models
  - Regression tree (RT), single neural networks (NN)

Configurations
- Number of test projects: N/6
- k : N/10

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Training</th>
<th>Test</th>
<th>Total</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td>NASA</td>
<td>50</td>
<td>10</td>
<td>60</td>
<td>6</td>
</tr>
<tr>
<td>NASA 93</td>
<td>78</td>
<td>15</td>
<td>93</td>
<td>9</td>
</tr>
<tr>
<td>USC</td>
<td>53</td>
<td>10</td>
<td>63</td>
<td>6</td>
</tr>
<tr>
<td>SDR</td>
<td>20</td>
<td>4</td>
<td>24</td>
<td>2</td>
</tr>
<tr>
<td>Desharnais</td>
<td>64</td>
<td>13</td>
<td>77</td>
<td>8</td>
</tr>
</tbody>
</table>
Validation design

Step 1. Data Preprocessing
- NASA Dataset
- NASA93 Dataset
- USC Dataset
- Desharnais Dataset
- SDR Dataset

(Preprocessed data)

Step 2. Applying effort estimation methods
- (1) RT
- (2) NN
- (3) ENN
- (4) ENNA

(25 times with random choosing)

Step 3. Measuring the estimation accuracy
Comparison and analysis of the effort estimation accuracy from each model

© KAIST SE LAB 2011
### Experimental results

<table>
<thead>
<tr>
<th></th>
<th>MMRE (%)</th>
<th></th>
<th></th>
<th>MME (%)</th>
<th></th>
<th></th>
<th>PRED(25) (%)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Best</td>
<td>Worst</td>
<td>Average (Std Dev)</td>
<td>Best</td>
<td>Worst</td>
<td>Average (Std Dev)</td>
<td>Best</td>
<td>Worst</td>
<td>Average (Std Dev)</td>
</tr>
<tr>
<td>RT</td>
<td>48.03</td>
<td>82.47</td>
<td>66.02 (11.64)</td>
<td>37.76</td>
<td>64.34</td>
<td>46.14 (7.99)</td>
<td>30.00</td>
<td>10.00</td>
<td>24.00 (8.43)</td>
</tr>
<tr>
<td>NASA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NN</td>
<td>122.14</td>
<td>385.29</td>
<td>184.46 (82.22)</td>
<td>89.63</td>
<td>118.75</td>
<td>97.97 (8.00)</td>
<td>20.00</td>
<td>0.00</td>
<td>5.00 (8.50)</td>
</tr>
<tr>
<td>ENN</td>
<td>49.23</td>
<td>66.13</td>
<td>55.71 (5.11)</td>
<td>32.50</td>
<td>47.39</td>
<td>41.82 (4.40)</td>
<td>40.00</td>
<td>30.00</td>
<td>32.00 (4.22)</td>
</tr>
<tr>
<td>ENNA</td>
<td>36.04</td>
<td>55.66</td>
<td>47.66 (6.62)</td>
<td>26.07</td>
<td>41.10</td>
<td>33.17 (4.34)</td>
<td>50.00</td>
<td>30.00</td>
<td>38.00 (6.33)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RT</td>
<td>334.97</td>
<td>532.68</td>
<td>394.34 (65.11)</td>
<td>51.71</td>
<td>67.10</td>
<td>58.94 (5.83)</td>
<td>26.67</td>
<td>6.67</td>
<td>20.67 (6.63)</td>
</tr>
<tr>
<td>NASA</td>
<td>232.13</td>
<td>1,925.40</td>
<td>527.69 (502.77)</td>
<td>94.19</td>
<td>197.69</td>
<td>107.76 (31.86)</td>
<td>13.33</td>
<td>0.00</td>
<td>6.00 (4.92)</td>
</tr>
<tr>
<td>NN</td>
<td>50.49</td>
<td>72.76</td>
<td>62.28 (7.56)</td>
<td>39.10</td>
<td>59.64</td>
<td>49.54 (8.31)</td>
<td>40.00</td>
<td>26.67</td>
<td>30.67 (4.66)</td>
</tr>
<tr>
<td>ENN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENNA</td>
<td>46.43</td>
<td>64.86</td>
<td>54.35 (7.03)</td>
<td>27.05</td>
<td>41.44</td>
<td>36.91 (4.19)</td>
<td>40.00</td>
<td>26.67</td>
<td>32.67 (3.78)</td>
</tr>
</tbody>
</table>
### Experimental results (cont’)

<table>
<thead>
<tr>
<th></th>
<th>USC</th>
<th>ENN</th>
<th>ENNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT</td>
<td>176.02</td>
<td>323.23</td>
<td>135.72</td>
</tr>
<tr>
<td></td>
<td>(217.49)</td>
<td>(75.20)</td>
<td>(3.16)</td>
</tr>
<tr>
<td>NN</td>
<td>1.934.80</td>
<td>5584.90</td>
<td>657.39</td>
</tr>
<tr>
<td></td>
<td>(4848.40)</td>
<td>(914.76)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>ENN</td>
<td>73.42</td>
<td>105.52</td>
<td>67.83</td>
</tr>
<tr>
<td></td>
<td>(20.46)</td>
<td>(8.80)</td>
<td>(4.22)</td>
</tr>
<tr>
<td>ENNA</td>
<td>56.99</td>
<td>85.44</td>
<td>61.42</td>
</tr>
<tr>
<td></td>
<td>(25.83)</td>
<td>(7.12)</td>
<td>(6.99)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>SDR</th>
<th>ENN</th>
<th>ENNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT</td>
<td>120.07</td>
<td>434.77</td>
<td>205.64</td>
</tr>
<tr>
<td></td>
<td>(241.59)</td>
<td>(266.72)</td>
<td>(12.91)</td>
</tr>
<tr>
<td>NN</td>
<td>315.17</td>
<td>2274.10</td>
<td>1892.00</td>
</tr>
<tr>
<td></td>
<td>(3791.30)</td>
<td>(5309.60)</td>
<td>(12.91)</td>
</tr>
<tr>
<td>ENN</td>
<td>36.25</td>
<td>49.85</td>
<td>38.60</td>
</tr>
<tr>
<td></td>
<td>(13.47)</td>
<td>(6.27)</td>
<td>(7.91)</td>
</tr>
<tr>
<td>ENNA</td>
<td>22.23</td>
<td>39.88</td>
<td>29.10</td>
</tr>
<tr>
<td></td>
<td>(10.95)</td>
<td>(11.92)</td>
<td>(15.81)</td>
</tr>
</tbody>
</table>
Experimental results (cont’)

<table>
<thead>
<tr>
<th></th>
<th>RT</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>68.07</td>
<td>251.80</td>
<td>119.36</td>
<td>39.74</td>
<td>62.73</td>
<td>47.67</td>
<td>30.77</td>
<td>15.39</td>
</tr>
<tr>
<td></td>
<td>(58.74)</td>
<td></td>
<td>(69.00)</td>
<td>(7.64)</td>
<td>(7.64)</td>
<td>(23.57)</td>
<td>(19.47)</td>
<td>(8.47)</td>
</tr>
<tr>
<td>NN</td>
<td>83.28</td>
<td>291.93</td>
<td>154.20</td>
<td>66.01</td>
<td>140.20</td>
<td>84.91</td>
<td>23.08</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(69.00)</td>
<td></td>
<td>(89.00)</td>
<td>(23.57)</td>
<td>(19.47)</td>
<td>(23.57)</td>
<td>(8.47)</td>
<td>(8.47)</td>
</tr>
<tr>
<td>ENN</td>
<td>52.52</td>
<td>64.36</td>
<td>57.85</td>
<td>47.40</td>
<td>53.64</td>
<td>50.06</td>
<td>30.77</td>
<td>15.39</td>
</tr>
<tr>
<td></td>
<td>(3.42)</td>
<td></td>
<td>(3.42)</td>
<td>(2.21)</td>
<td>(2.21)</td>
<td>(2.21)</td>
<td>(5.13)</td>
<td>(5.13)</td>
</tr>
<tr>
<td>ENNA</td>
<td>42.05</td>
<td>55.76</td>
<td>49.82</td>
<td>35.59</td>
<td>49.03</td>
<td>44.13</td>
<td>46.15</td>
<td>23.08</td>
</tr>
<tr>
<td></td>
<td>(5.05)</td>
<td></td>
<td>(5.05)</td>
<td>(4.57)</td>
<td>(4.57)</td>
<td>(4.57)</td>
<td>(7.30)</td>
<td>(7.30)</td>
</tr>
</tbody>
</table>

- ENN outperforms over RT and NN
- ENNA outperforms over all other methods
- ENN, ENNA are stable model
  - Because they have small standard deviation
Conclusion

 ItemType

 Contributions
- Propose ensemble of neural network with associative memory
  - It makes accurate estimations with small data
  - It generates stable results
- Improve the accuracy of effort estimation
  - Over regression tree and single neural networks

 Future work
- Use RT model instead of NN
- Analyze the sensitivity of model to the size of dataset
Limitations

- There is no evidence for using the number of test and k
  - # of test projects: N/6
  - # of similar projects: N/10
- Inherent difference between projects is not considered when bias is calculated
Thank You.