Running CCLearner with BigCloneBench

Liuqing Li, He Feng, Wenjie Zhuang, Na Meng and Barbara Ryder
{lruqing, fenghe, kaito, nm8247}@vt.edu, ryder@cs.vt.edu

Department of Computer Science
Virginia Polytechnic Institute and State University
Blacksburg, VA, USA
September 21, 2017
Code Clone

Developers copy and paste code to improve programming productivity

Clone detections tools are needed to help bug fixes or program changes
Methodology

Code Clone Detection Problem

Classification Problem
Approach

Source code

Method Extraction

Method Pair Enumerator

Clone Pairs

Non-Clone Pairs

Training

Deep Learning

Classifier

Testing

Clone Pairs

Non-Clone Pairs
BigCloneBench*

- **URL**
  - [https://github.com/clonebench/BigCloneBench](https://github.com/clonebench/BigCloneBench)

- **Benchmark Database**

- **Java Codebase**
  - **IJaDataset 2.0 – Reduced Version**
  - **10 folders**
    - Folder #4 – Training
    - Others – Testing

---

GitHub – Source Code
https://github.com/liuqingli/CCLearner

VirtualBox – Runnable Environment
https://goo.gl/k6rjDn
Novice Users

- `java -jar CCLearner_Feature.jar`
- `java -jar CCLearner_Train.jar`
- `java -jar CCLearner_Test.jar`
## Expert Users

### Feature

<table>
<thead>
<tr>
<th>Feature</th>
<th>number</th>
<th>name</th>
</tr>
</thead>
</table>

### Training

<table>
<thead>
<tr>
<th>Training</th>
<th>number of iterations</th>
<th>number of input nodes</th>
<th>number of hidden nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>… … …</td>
<td>… … …</td>
<td>… … …</td>
<td>… … …</td>
</tr>
</tbody>
</table>
Thank you!

Research Talk

CC Learner: A Deep Learning-Based Clone Detection Approach

11:40 - 12:00, Friday, Grand Ballroom B