<table>
<tr><th rowspan="2">cases</th><th>doc_1</th><th>doc_2</th><th>decision</th><th>id</th></tr>
<tr><td><table>
<tr><th>authors</th><td><ul><li>Michael Weiss</li><li>Rwiddhi Chakraborty</li><li>Paolo Tonella</li></ul></td></tr>
<tr><th>title</th><td>A Review and Refinement of Surprise Adequacy</td></tr>
<tr><th>publication_date</th><td>2021-06-01 00:00:00</td></tr>
<tr><th>source</th><td>SupportedSources.OPENALEX</td></tr>
<tr><th>journal</th><td>2021 IEEE/ACM Third International Workshop on Deep Learning for Testing and Testing for Deep Learning (DeepTest)</td></tr>
<tr><th>volume</th><td></td></tr>
<tr><th>doi</th><td>10.1109/deeptest52559.2021.00009</td></tr>
<tr><th>urls</th><td><ul><li>https://openalex.org/W3185950485</li><li>https://doi.org/10.1109/deeptest52559.2021.00009</li><li>http://arxiv.org/pdf/2103.05939</li></ul></td></tr>
<tr><th>id</th><td>id-8770103532469784660</td></tr>
<tr><th>abstract</th><td></td></tr>
<tr><th>versions</th><td></td></tr>
</table></td><td><table>
<tr><th>authors</th><td><ul><li>Paolo Tonella</li><li>Rwiddhi Chakraborty</li><li>Michael Weiss</li></ul></td></tr>
<tr><th>title</th><td>A Review and Refinement of Surprise Adequacy</td></tr>
<tr><th>publication_date</th><td>2021-03-10 00:00:00</td></tr>
<tr><th>source</th><td>SupportedSources.PAPERS_WITH_CODE</td></tr>
<tr><th>journal</th><td></td></tr>
<tr><th>volume</th><td></td></tr>
<tr><th>doi</th><td></td></tr>
<tr><th>urls</th><td><ul><li>https://arxiv.org/pdf/2103.05939v1.pdf</li><li>https://github.com/coinse/sadl</li></ul></td></tr>
<tr><th>id</th><td>id-1048397052880443473</td></tr>
<tr><th>abstract</th><td>Surprise Adequacy (SA) is one of the emerging and most promising adequacy criteria for Deep Learning (DL) testing. As an adequacy criterion, it has been used to assess the strength of DL test suites. In addition, it has also been used to find inputs to a Deep Neural Network (DNN) which were not sufficiently represented in the training data, or to select samples for DNN retraining. However, computation of the SA metric for a test suite can be prohibitively expensive, as it involves a quadratic number of distance calculations. Hence, we developed and released a performance-optimized, but functionally equivalent, implementation of SA, reducing the evaluation time by up to 97\%. We also propose refined variants of the SA omputation algorithm, aiming to further increase the evaluation speed. We then performed an empirical study on MNIST, focused on the out-of-distribution detection capabilities of SA, which allowed us to reproduce parts of the results presented when SA was first released. The experiments show that our refined variants are substantially faster than plain SA, while producing comparable outcomes. Our experimental results exposed also an overlooked issue of SA: it can be highly sensitive to the non-determinism associated with the DNN training procedure.</td></tr>
<tr><th>versions</th><td></td></tr>
</table></td><td>DUPLICATES</td><td>118</td></tr>
</table>