Likelihood Ratios For Out-of-Distribution Detection

Jie Ren*, Peter J. Liu, Emily Fertig, Jasper Snoek, Ryan Poplin, Mark A. DePristo, Joshua V. Dillon, Balaji Lakshminarayanan*
Motivation: Why is OOD detection important?

- Bacteria identification based on genomic sequences
  - ACGTTAACCAACC...GGCTTC → label
  - Holds the promise of early detection of disease
- Classifier can achieve high accuracy on cross-validation
- But, the classifier can perform poorly in real world:
  - 60-80% data belonging to as yet unknown bacteria
  - Assign high-confidence predictions to OOD inputs, than say “I don’t know”
- Need accurate OOD detection to ensure safe deployment of classifier
Generative models for OOD detection?

- **Pros**: do not require labeled data; model the input distribution $p(x)$ and then evaluate the likelihood of new inputs

- **Cons**: can **assign higher likelihood to OOD** inputs!

![Density vs Log-likelihood](image1.png)

**Fashion-MNIST (in-dist.) vs. MNIST (OOD)**

![Histogram vs Log-likelihood](image2.png)

**Genomics**
What does $p(x)$ represent?

- **Examples of Background vs. Semantics:**
  - Images: background + objects
  - Text: stop words + key words
  - Genomics: GC background + motifs
  - Speech: background noise + speaker

- Likelihood $p(x)$ has to explain both semantic and background components

$$p(x) = p(x_B) p(x_S)$$

can be dominant

the focus

- Humans ignore background and focus primarily on semantics for OOD
- **Question:** how do we automatically extract semantic component of $p(x)$?
Likelihood Ratio for OOD Detection

To focus on $x_S$ we propose:

1. Training a **background model** on perturbed inputs
2. Computing the likelihood ratio

$$LLR(x) = \log \frac{p_\theta(x)}{p_{\theta_0}(x)} = \log \frac{p_\theta(x_B) \cdot p_\theta(x_S)}{p_{\theta_0}(x_B) \cdot p_{\theta_0}(x_S)} \approx \log \frac{p_\theta(x_S)}{p_{\theta_0}(x_S)}$$

- LLR is a **background contrastive score**: the significance of the semantics compared with the background.
Which pixels contribute the most to likelihood (ratio)?

- PixelCNN++ model trained on FashionMNIST
- Heatmap showing per-pixel contributions on Fashion-MNIST (in-dist) and MNIST (OOD)

\[
\log p_\theta(x_d|x_{<d})
\]

\[
\log p_\theta(x_d|x_{<d}) - \log p_{\theta_0}(x_d|x_{<d})
\]

Likelihood is dominated by background pixels, which explains why MNIST (OOD) is assigned higher \( p(x) \)

Likelihood ratio focuses more on the semantic pixels and significantly outperforms likelihood on OOD detection
OOD detection for genomic sequences

<table>
<thead>
<tr>
<th>Method</th>
<th>AUROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood</td>
<td>0.626</td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>0.755</td>
</tr>
<tr>
<td>Classifier-based $p(y</td>
<td>x)$</td>
</tr>
<tr>
<td>Classifier-based Entropy</td>
<td>0.634</td>
</tr>
<tr>
<td>Classifier-based ODIN</td>
<td>0.697</td>
</tr>
<tr>
<td>Classifier Ensemble 5</td>
<td>0.682</td>
</tr>
<tr>
<td>Classifier-based Mahalanobis Distance</td>
<td>0.525</td>
</tr>
</tbody>
</table>

Likelihood is heavily affected by GC bias

Likelihood Ratio corrects for GC bias
Summary

- Likelihood from deep generative models can be affected by background
- The proposed Likelihood Ratio method effectively corrects for background, and outperforms the raw likelihood on OOD detection

- Release a realistic benchmark dataset for OOD detection in genomics
- Our method achieves SOTA performance on genomic dataset

New benchmark dataset + code is available at https://github.com/google-research/google-research/tree/master/genomics_ood