Machine Learning & Deep Learning in Quantitative Image Analysis: An Introduction and Primer

David Egan & Wienand Omta, Core Life Analytics
Juan Caicado, Broad Institute of MIT & Harvard
Oren Kraus & Sam Cooper, Phenomic AI
Overview

1. Introduction to Machine Learning; David Egan & Wienand Omta
2. Deep Learning: What's All The Fuss About; Sam Cooper
3. Successes & Failures; Oren Kraus
4. Interpretability of Deep Learning Models; Juan Caicado
1.

Introduction to Machine Learning
Artificial Intelligence, Machine Learning, Deep Learning

Artificial Intelligence
Techniques that enable computers to mimic human intelligence

Deep Learning
Subset of ML that uses multilayered neural networks.

Machine Learning
Uses statistical techniques that enables machines to improve at tasks with experience.
Unsupervised Methods
The data set is unlabelled. Looking for hidden structure in the data in an unbiased way e.g. clustering
Supervised Methods
You use a known annotated or “labeled” data set to help you classify items in an unknown “unlabeled” data set.

Data Set

Unsupervised vs Supervised Methods

Unlabeled Data

Labeled data

Unlabeled Data

Labeled Data

Model

Classify

Label A

Label B
Machine Learning

- Train, and test a model
- Use the model for new data sets
- Should be useful for generalization, prediction
- “Allows us to solve seemingly impossible tasks”
- Show it lots of examples and let the software do the rest.

Build Model With Labeled Data

Apply Model to New Unlabeled Data
How Does ML Build a Model

- Classic example: A Linear Model
- $y = mx + c$
- In ML $y' = b + w_1x_1$
- $y'$ label, $b$ bias, $x_1$ feature, $w_1$ weight
- Many Features: $y' = b + w_1x_1 + w_2x_2 + w_3x_3 + w_3x_3$
- $y'$ = cancerous cell, $x_{1-n}$ features extracted from image
How Does ML Build a Model

Apply Model

Measure Loss

Modify $w_1$ and $b$

- Minimize the loss
- Iterative fashion
- Get the best possible model as efficiently as possible
- Playground Exercise

$$y' = b + w_1 x_1$$
How Does ML Build a Model

Compute Parameter Updates

Features

Model
(Prediction function)

Label

Compute Loss
Avoiding Overtraining

Generalization

- Should be able to use your model for new data sets
- For complex data sets your model may be too specific to your training set; Overtraining
- Need to test your model against another random subset of your data, the test set
Training Set & Test Set

- Random Split Into Training Set & Test Set, e.g. 80/20
- Does not overlap with Training Set
- Report Accuracy of Your Model

Model

Data Set

Accuracy

Unlabeled Data

Labeled data

Classify

Unlabeled Data

Training Set

Test Set

Label A

Label B
Model Accuracy

True Positive Rate = TP/TP + FN

True Negative Rate = TN/TN + FN

- At very low classification threshold, high TP, high FP
- At very high classification threshold, low TP, low FP
- In a good model the line is very steep
- AUC is very high, approaches 1
- Other Metrics; F1 and Kappa
Example

Unsupervised Method: Distance Score
Example

Unsupervised Method: Clustering
Example

Supervised Method: Random Forest ML
Statistics vs. Machine Learning

- Traditional Statistics; Use a sample to represent a much larger data set.
- Was very useful when it was impossible to gather information on the whole data set.
- With computing power and data bases statistical methods can now be used on complete data sets
- ML uses many of the same methods
Deep Learning: What’s All The Fuss About?
Most classification techniques over the last 20-30 years are comparable... then deep-learning arrived

1980’s: Multilayer Perceptrons ->
Backpropagation solved by Rumelhart, Hinton, Williams, LeCun

1990’s: Random forest classifiers, Support Vector Machines, lots of other techniques developed.

1998: AlexNet (LeCun et.al.) one of the first CNN’s, outperforms all existing machine learning approaches on MNIST

2000’s: GPU’s become cheap

2010 -> today: Massive ‘ImageNet’ advances by CNN’s launches deep-learning explosion

Table 1: Comparison of results on ILSVRC-2010 test set. In *italics* are best results achieved by others.

<table>
<thead>
<tr>
<th>Model</th>
<th>Top-1</th>
<th>Top-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparse coding [2]</td>
<td>47.1%</td>
<td>28.2%</td>
</tr>
<tr>
<td>SIFT + FVs [24]</td>
<td>45.7%</td>
<td>25.7%</td>
</tr>
<tr>
<td>CNN</td>
<td>37.5%</td>
<td>17.0%</td>
</tr>
</tbody>
</table>

Figure 3: 96 convolutional kernels of size $11 \times 11 \times 3$ learned by the first convolutional layer on the $224 \times 224 \times 3$ input images. The top 48 kernels were learned on GPU 1 while the bottom 48 kernels were learned on GPU 2. See Section 6.1 for details.

Advantages

1. For most machine learning tasks to date, empirically have just worked better than anything else.

2. Features and pipelines don’t need to be re-engineered for each new data set -> faster analysis times.

3. Easy to produce results that can be validated: 
   1. Qualitative
   2. Retrospective prediction
   3. Prospective prediction

Logistic Regression / Linear Perceptron

\[\sum f(x) \rightarrow \text{Probability of 5 = 0.1}\]

\[\sum f(x) \rightarrow \text{Probability of 6 = 0.2}\]

\[\sum f(x) \rightarrow \text{Probability of 7 = 0.1}\]

\[\sum f(x) \rightarrow \text{Probability of 8 = 0.9}\]

\[\sum f(x) \rightarrow \text{Probability of 9 = 0.2}\]
Multiple layers have a higher IQ

A single layer can solve this challenge

But not this one
Multiple layers have a higher IQ

**A single layer can solve this challenge**

A single layer can solve this challenge.

**But not this one**

But not this one.
Multilayer Perceptron

By stacking lot’s of non linear-layers on top of each other [Hidden layers] we can learn to spot complex patterns

MNIST TEST ACCURACY 97 %
The visual cortex is ‘locally receptive’
Convolutional networks also have a receptive field.

Weights are 2D ‘Kernels’ -> These learn features

Figure 3: 96 convolutional kernels of size $11 \times 11 \times 3$ learned by the first convolutional layer on the $224 \times 224 \times 3$ input images. The top 48 kernels were learned on GPU 1 while the bottom 48 kernels were learned on GPU 2. See Section 6.1 for details.
Deep convolutional neural networks

Stacking convolutional layers allows learning of abstract features

Going deeper allows us to classify better

**ILSVRC top-5 error on ImageNet**

<table>
<thead>
<tr>
<th>Year</th>
<th>Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>30</td>
</tr>
<tr>
<td>2011</td>
<td>22.5</td>
</tr>
<tr>
<td>2012</td>
<td>15</td>
</tr>
<tr>
<td>2013</td>
<td>7.5</td>
</tr>
<tr>
<td>2014</td>
<td>0</td>
</tr>
<tr>
<td>Human</td>
<td>0</td>
</tr>
<tr>
<td>ArXiv 2015</td>
<td>0</td>
</tr>
</tbody>
</table>

**GoogLeNet, 2014 ImageNet competition winner**
Many architectures and examples go beyond classification:

- **Fully convolutional nets**, $\rightarrow$ **Segmentation**
- **Graph convolution**, $\rightarrow$ **Molecules**
- **Deep reinforcement learning**, $\rightarrow$ **Actions**
- **Recurrent neural networks**, $\rightarrow$ **DNA**
Where next?

![Graph showing the progression of intelligence over time, with labels for various levels of intelligence including EINSTEIN, DUMB HUMAN, CHIMP, BIRD, ANT, and AI INTELLIGENCE. The graph indicates an exponential increase in intelligence with the label "Reality." The text "Haha that's adorable, the funny robot can do monkey tricks!" is noted on the graph.]
3. Successes & Failures

A. HEALTHY

B. DISEASED

- Hemorrhages

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlexNet</td>
<td>43.45</td>
</tr>
<tr>
<td>VGG-19</td>
<td>25.76</td>
</tr>
<tr>
<td>Net-101</td>
<td>22.63</td>
</tr>
<tr>
<td>inception v3</td>
<td>22.55</td>
</tr>
<tr>
<td>net-161</td>
<td>22.35</td>
</tr>
</tbody>
</table>
Deep Learning in High Content Screening – Community

• Field really picked up steam in 2016 with the first CytoData meeting at the Broad Institute

CytoData

Data-analysis strategies for image-based cell profiling
– Nat Meth 2017
Deep Learning in Microscopy – Early Work

Yeast Proteome Dynamics from Single Cell Imaging and Automated Analysis – Cell 2015

- > **70,000** individual cell annotated
- First quantitative assessment of abundance and localization of 4,100 proteins
- Classifier: ensemble of 60 binary SVMs
Limitations with Existing Workflows

- Challenges: Difficult to reproduce, some classes hard to classify
- Challenges: Difficult to optimize experimental protocol & assay together
Supervised Deep Learning for Protein Localization

Automated analysis of high-content microscopy data with deep learning – Molecular Systems Biology 2016

- Applied deep learning to classify protein localization
- Significantly outperformed previous SVM benchmark
- Much simpler workflow – single CNN based classifier
- Transferred to other datasets easily
Supervised Deep Learning for Protein Localization

Single Cell Classification

Transfer Learning

Benefits
- Better accuracy
- Easily transferable to similar datasets

Limitations
- Still trained on a significant number of annotated single cells
- Workflow can’t be easily reused for new screens
- Only supervised workflow
Challenge: Classifying Images with Mixed Populations

Bottleneck: Most analyses based on labeling single cells

Can we do Better?
Deep Multiple Instance Learning

Classifying and segmenting microscopy images with deep multiple instance learning – Bioinformatics 2016

- Technique for training deep learning models without segmentation
- Aggregates across single cell objects without requiring single cell labels
Deep Multiple Instance Learning - Workflows

**Screening workflow**
- Training on controls
- Score entire screen

**Dose Response workflow**
- Train on low/high dose
- Evaluated Dose Response

---

### Screening workflow

- **Training on controls**
  - +ve control
  - -ve control

- **Score entire screen**
  - Image showing scored screen

### Dose Response workflow

- **Train on low/high dose**
  - -ve control
  - +ve control

- **Evaluated Dose Response**
  - Graph showing dose response
  - Log [µM] vs Multiple Instance Learning DR

---

EC50 [µM] Z'
MIL PE MIL PE
HeLa 10.4 7.4 0.92 0.79
HCT116 11.3 6.8 0.94 0.46
PANC 6.2 6.3 0.81 0.82
Deep Learning **Failure** (Until Recently)

**Screening Workflow**
- Score treatments based on control wells
  - Clear way to find hits - MIL works well
  - Biased by controls - can’t identify/relate new phenotypes

**Profiling Workflow**
- No controls to train on
  - Unbiased, find new phenotypes, relate conditions to each other
  - Difficult to validate quantitatively - results explored visually
Phenotypic Profiling - by Autoencoders

- In profiling workflows, images need to be represented in such a way that they can be compared numerically.
- Deep learning approaches struggle with this task as they rely on lots of annotated examples to learn how to represent images – Not available.
- Traditional features do not require annotations and can be computed directly.
- Autoencoders, a popular deep learning approach for unsupervised feature learning, doesn’t work well with HCS data – just learns most obvious variations.

Weakly supervised learning of single-cell feature embeddings – CVPR 2018

• More recent approach - Train model to predict treatments to learn embedding
  + Unbiased - no annotations required (can be based on well ID)
  + Model forced to learn features that discriminate between treatments
  - Doesn’t scale well to large screens with >100,000 compounds
  - Many treatments have identical [no] effects - difficult optimization
Phenotypic Profiling – by Weakly Supervised Learning

Weakly Supervised Features

Extracted Features
**Phenotypic Profiling – by Cell Inpainting**

*Learning unsupervised feature representations for single cell microscopy images with paired cell inpainting – bioarxiv 2018*

- More recent approach - Train model to ‘paint’ probe channel of target cell, given source cell and context channel of target cell
  - Unbiased - no annotations required - learns nice feature representations
  - Overcome scalability issue of training on treatments
    - Some patterns cannot be predicted from context channel
    - Single cell heterogeneity in treatments - some pairs don’t match
4. Interpretability of Deep Learning Models
Deep learning models are black boxes

- Features do not have names
- Decisions can’t be explained easily

But, what is interpretability?

- Naming features
- Explaining decisions
- Finding example images
- Supporting new hypotheses

Another perspective on interpretation
1. ANALYZE OUTPUTS AND ERROR MODES

There are many properties of the data that can be investigated using better image processing models.
Consider Object Segmentation

Segmentation

Otsu’s thresholding method
What objects are hard to segment?

Segmentation errors by object size

<table>
<thead>
<tr>
<th>Area (pixels)</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large</td>
<td>901</td>
<td>&gt;</td>
</tr>
<tr>
<td>Medium</td>
<td>626</td>
<td>900</td>
</tr>
<tr>
<td>Small</td>
<td>251</td>
<td>625</td>
</tr>
<tr>
<td>Tiny</td>
<td>0</td>
<td>250</td>
</tr>
</tbody>
</table>

Caicedo et al. bioRxiv 335216, 2018
Why some image types are more difficult to segment?

Data Science Bowl 2018: Nucleus Segmentation Challenge
Other examples outside imaging

Geology researchers improve earthquake theory by analyzing the outputs of a deep learning predictor.

2. SIMILARITIES REVEAL PATTERNS

Single readouts may not tell the full story. Look at relationships between data points using similarity metrics.
From multi-dimensional features to similarities
Morphology-based analysis of single cells

EGFR_p.T790M, p.L858R.o

Control  Wild Type  Mutant
Morphology analysis of single cells

Variant impact: 64.0%
4. CORRELATIONS WITH OTHER VARIABLES

Image-based features may predict other biological readouts very accurately.
Predicting the readout of biological assays

Not all assays can be predicted, which aids interpretation of image-based features.

5. GENERATING EXAMPLE IMAGES

Understanding morphological variation by finding example images that explain important changes
Generating example images

Low dose

High dose

Treatments

Goldsborough, et al. NIPS MLCB 2017
Conclusions

• Interpretation of models is beyond feature names
• Better predictors can enable better interpretations
• 5 examples of how to interpret biological image data using deep learning