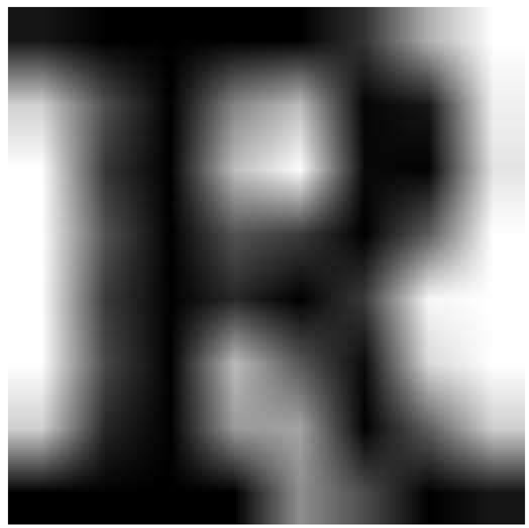
Optical Character Recognition Domain Expert Approximation Through Oracle Learning

Joshua Menke
NNML Lab
BYU CS
josh@cs.byu.edu

March 24, 2004

Optical Character Recognition (OCR)

- *optical character recognition* (OCR): given an image, give the letter

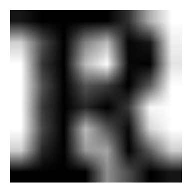


↓
R

OCR with ANNs

Artificial Neural Networks (ANNs)

- Powerful adaptive machine learning models
- Trained for OCR to recognize images as letters
- 98%+ accuracy



$\downarrow$

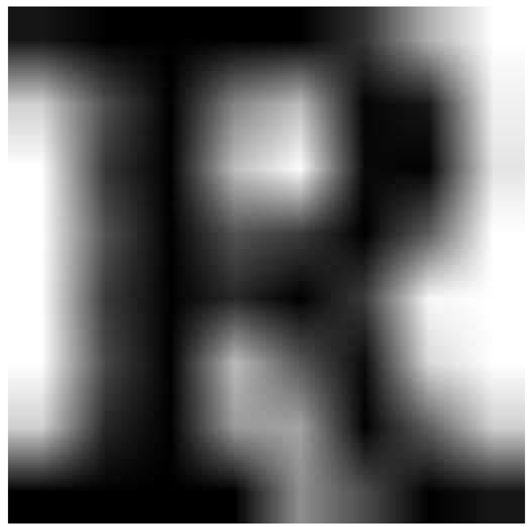
$\boxed{\text{ANN}}$

$\downarrow$

R

Problem: Varying Noise

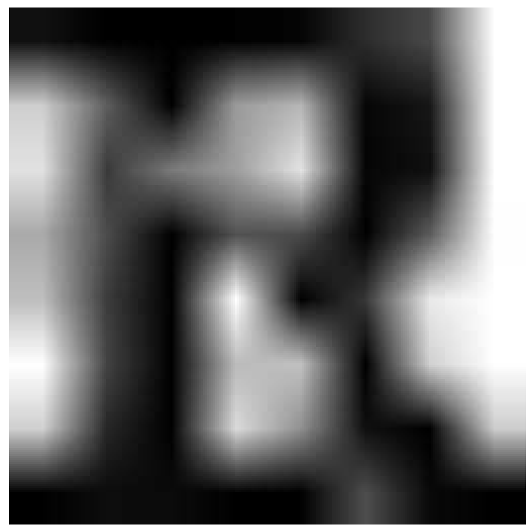
The amount of noise in a given image can vary for the same letter



Yields two domains, noisy and clean.

Problem: Varying Noise

The amount of noise in a given image can vary for the same letter



Yields two domains, noisy and clean.

Varying Noise: Common Solution

- Train one ANN (ANN_{mixed}) on clean and noisy images mixed
- Problem: Noisy regions in the domain are more difficult to approximate
  - ANNs will learn the easier, clean images first.
  - Then will continue training to learn the noisy regions
  - The ANN can overfit the clean domain, lowering overall accuracy

Domain Experts

- The Domain Experts:
  - ANN_{clean} trains on / recognizes clean images
  - ANN_{noisy} trains on / recognizes noisy images

- Separates clean and noisy training, so no overfit to clean images.

- Problem: Choosing the right ANN given a new letter.

  Solutions*:

  - Train a separate ANN to distinguish clean from noisy letters.
  - Use both ANNs and choose the one with the most confidence.

    *Difficult to do in practice

The Oracle Learning Process

Originally used to create reduced sized ANNs.

1. Obtain the Oracle: Large

2. Label the Data

3. Train the *Oracle-Trained Network* (OTN): Small

The Oracle Learning Process

Obtain the most accurate ANN regardless of size.

$$\begin{array}{c} \boxed{ANN_{large}} \\ \uparrow \\ \boxed{\text{Training Data}} \end{array}$$

The Oracle Learning Process

Use the trained oracle to relabel the training data with its own outputs.

$$\begin{array}{c}\boxed{\text{Relabeled Training Data}}\\ \uparrow \\ \boxed{ANN_{large}} \\ \uparrow \\ \boxed{\text{Training Data}}\end{array}$$

The Oracle Learning Process

Use the relabeled training set to train a simpler ANN.

Oracle Outputs = New Targets

$\uparrow$

ANN_{small}

$\uparrow$

Oracle-labeled Training Data

Domain Expert Approximation Through Oracle Learning: *Bestnets*

- We introduce the *bestnets* method.
- Use Oracle learning [7] to train an ANN to approximate the behavior of:
  - ANN_{clean} on clean images
  - ANN_{noisy} on noisy images
- Successfull approximation gives $ANN_{bestnets}$:
  - The accuracy of ANN_{clean} on clean images
  - The accuracy of ANN_{noisy} on noisy images
  - An implicit ability to distinguish between clean and noisy
  - No fear of overfitting. Overfitting the oracles is desirable.

Prior Work

- Approximation
  - Menke et al. [7, 6]: Oracle Learning
  - Domingos [5]: Approximated a bagging [1] ensemble with decision trees [8]
  - Zeng and Martinez [9] approximated a bagging ensemble with an ANN
  - Craven and Shavlik approximated an ANN with rules [3] and trees [4]
  - Bestnets approximates domain experts (novel)

- Varying Noise: Mostly unrelated work.
  - Assume one type of noise OR
  - Vary the noise but train / test each separately OR
  - Assume knowledge about the type of noise (SNR, etc.)
  - Not always realistic

Bestnets Method for OCR

Three steps:

1. Obtain the Oracles. In this case two oracles:
   - Find the best ANN for clean only images (ANN_{clean})
   - Find the best ANN for noisy only images (ANN_{noisy})
2. Relabel the images with the oracles
   - Relabelel clean images with ANN_{clean}'s outputs
   - Relabelel noisy images with ANN_{noisy}'s outputs
3. Train a single ANN ($ANN_{bestnets}$) with the relabeled images

Note About Output Targets

The OCR ANNs have an output for every letter we'd like to recognize. Given an image, the output corresponding to the correct letter should have a higher value than the other outputs. These values range between 0 and 1.

To train an ANN to do this every incorrect output is trained to output 0 and the correct one 1.

With Oracle Learning, instead of training to 0-1, the OTN trains to output what its oracles output instead, always more relaxed (greater than 0 or less than 1).

May be an easier to learn according to Caruana [2].

Bestnets Process

Train the domain experts.

$$
\begin{array}{ccc}
\boxed{ANN_{noisy}} & & \boxed{ANN_{clean}} \\
\uparrow & & \uparrow \\
\boxed{\text{Noisy Training Images}} & & \boxed{\text{Clean Training Images}}
\end{array}
$$

Bestnets Process

Use the trained experts to relabel the training data with their own outputs.

$$\begin{tikzcd}
\boxed{\text{Relabeled Noisy Images}} & \boxed{\text{Relabeled Clean Images}} \\
\boxed{ANN_{noisy}} \arrow[u] & \boxed{ANN_{clean}} \arrow[u] \\
\boxed{\text{Noisy Training Images}} \arrow[u] & \boxed{\text{Clean Training Images}} \arrow[u]
\end{tikzcd}$$

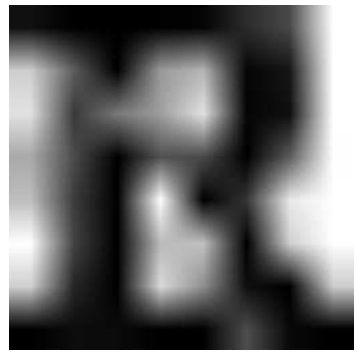
Bestnets Process

Use the relabeled training set to train a single ANN on the oracles' outputs.

$$
\begin{array}{c}
\boxed{\text{Expert Outputs} = \text{New Targets}} \\
\uparrow \\
\boxed{ANN_{bestnets}} \\
\uparrow \\
\boxed{\text{Relabeled Clean and Noisy Training Images}}
\end{array}
$$

Example: Original Training Image

Image



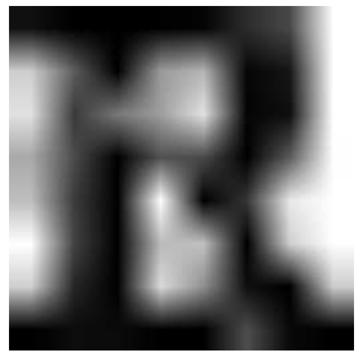
Target

All 0's except for the output corresponding to R which is 1

Domain

Noisy

Example: Getting the Oracle's Outputs



$$\downarrow$$

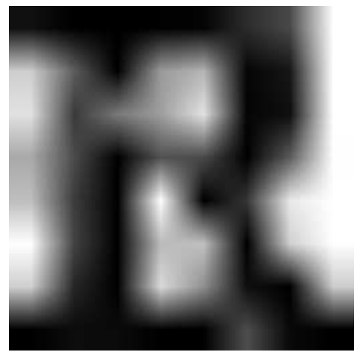
$$\boxed{ANN_{noisy}}$$

$$\downarrow$$

$$< 0.2, 0.3, 0.13, ..., R = 0.77, ..., 0.44 >$$

Example: Resulting Training Image

Image



Target

$$< 0.2, 0.3, 0.13, ..., R = 0.77, ..., 0.44 >$$

Experiment

1. Train ANN_{clean} on only the clean images
2. Train ANN_{noisy} on only the noisy images
3. Relabel the clean letter set's output targets with ANN_{clean}'s outputs
4. Relabel the noisy letter set's output targets with ANN_{noisy}'s outputs
5. Train a single ANN ($ANN_{bestnets}$) on the relabeled images from both sets
6. Train standard ANN_{mixed} on both clean and noisy with standard 0-1 targets

Initial Results

ANN1	ANN2	Data set	Difference	p-value
ANN_{clean}	ANN_{mixed}	Clean	0.0307	< 0.0001
ANN_{noisy}	ANN_{mixed}	Noisy	0.0092	< 0.0001
$ANN_{bestnets}$	ANN_{mixed}	Mixed	0.0056	< 0.0001
ANN_{clean}	$ANN_{bestnets}$	Clean	0.0298	< 0.0001
ANN_{noisy}	$ANN_{bestnets}$	Noisy	-0.0011	0.1607

p-values from a McNemar test comparing the two classifiers in each row on a test set.

Conclusion and Future Work

- Conclusion:

  The bestnets-trained ANN:

  - Improves over standard (mixed) training
  - Retains the performance of ANN_{noisy}

- Future Work

  - Increase the improvement focusing on clean
  - Investigate why it works (Caruana [2], may be easier to learn)

References

[1] L. Breiman. Bagging predictors. *Machine Learning.*, 24(2):123–140, 1996.

[2] Rich Caruana, Shumeet Baluja, and Tom Mitchell. Using the future to "sort out" the present: Rankprop and multitask learning for medical risk evaluation. In David S. Touretzky, Michael C. Mozer, and Michael E. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8, pages 959–965, Cambridge, MA, 1996. The MIT Press.

[3] Mark Craven and Jude W. Shavlik. Learning symbolic rules using artificial neural networks. In Paul E. Utgoff, editor, *Proceedings of the Tenth International Conference on Machine Learning*, pages 73–80, San Mateo, CA, 1993. Morgan Kaufmann.

[4] Mark W. Craven and Jude W. Shavlik. Extracting tree-structured representations of trained networks. In David S. Touretzky, Michael C. Mozer, and Michael E. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8, pages 24–30, Cambridge, MA, 1996. The MIT Press.

[5] Pedro Domingos. Knowledge acquisition from examples via multiple models. In *Proceedings of the Fourteenth International Conference on Machine Learning*, pages 98–106, San Francisco, 1997. Morgan Kaufmann.

[6] Joshua Menke and Tony R. Martinez. Simplifying ocr neural network through oracle learning. In *Proceedings of the 2003 International Workshop on Soft Computing Techniques in Instrumentation, Measurement, and Related Applications*. IEEE Press, 2003.

[7] Joshua Menke, Adam Peterson, Michael E. Rimer, and Tony R. Martinez. Neural network simplification through oracle learning. In *Proceedings of the IEEE International Joint Conference on Neural Networks IJCNN'02*, pages 2482–2497. IEEE Press, 2002.

[8] J.R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA, 1993.

[9] Xinchuan Zeng and Tony Martinez. Using a neural networks to approximate an ensemble of classifiers. *Neural Processing Letters.*, 12(3):225–237, 2000.