The Open Set Recognition Problem

Walter J. Scheirer
School of Engineering and Applied Sciences,
Department of Cellular and Molecular Biology, and
Center for Brain Science, Harvard University
Benchmarks in computer vision

Assume we have examples from all classes:

- airplanes
- elephant
- soccer ball
- car
- water lily

Caltech 256
Out in the real world…

Detect the cars in this image

while rejecting the trees, signs, telephone poles…

“All positive examples are alike; each negative example is negative in its own way”

Zhao and Huang (with some help from Tolstoy)
CVPR 2001
What is the general object recognition problem?

• Duin and Pekalska*: how one should approach multi-class recognition is still an open issue
  - Is it a series of binary classifications?
  - Is it a search performed for each possible class?
  - What happens when some classes are ill-sampled, not sampled at all or undefined?

Vision problems in order of "openness"

Multi-class Classification
- Training and testing samples come from known classes

Face Verification
- Claimed identity, possibility for impostors

Detection
- One class, everything else in the world is negative

Open Set Recognition
- Multiple known classes, many unknown classes

Let’s formalize openness

\[
\text{openness} = 1 - \sqrt{\frac{2 \times |\text{training classes}|}{|\text{testing classes}| + |\text{target classes}|}}
\]
Examples of openness values

<table>
<thead>
<tr>
<th></th>
<th>Targets</th>
<th>Training</th>
<th>Testing</th>
<th>Openness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Typical Multi-class</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>0%</td>
</tr>
<tr>
<td><strong>Face Verification</strong></td>
<td>12</td>
<td>12</td>
<td>50</td>
<td>38%</td>
</tr>
<tr>
<td>Typical Detection</td>
<td>1</td>
<td>100,000</td>
<td>1,000,000</td>
<td>55%</td>
</tr>
<tr>
<td><strong>Object Recognition</strong></td>
<td>88</td>
<td>12</td>
<td>88</td>
<td>63%</td>
</tr>
<tr>
<td><strong>Object Recognition</strong></td>
<td>88</td>
<td>6</td>
<td>88</td>
<td>74%</td>
</tr>
<tr>
<td><strong>Object Recognition</strong></td>
<td>212</td>
<td>6</td>
<td>212</td>
<td>83%</td>
</tr>
</tbody>
</table>
Fundamental multi-class recognition problem

\[
\arg\min_f \left\{ R_\mathcal{I}(f) := \int_{\mathbb{R}^d \times \mathbb{N}} L(x, y, f(x)) P(x, y) \right\}
\]

Ideal Risk \hspace{1cm} Loss Function \hspace{1cm} Joint Distribution

Undefined for open set recognition!

Open Space

Negatives

Positives

Specialization
Open Space

- Open space is the space far from known data
- We need to address the infinite half-space problem of linear classifiers
- Principle of Indifference*
  - If there is no known reason to assign probability, alternatives should be given equal probability
  - One problem: we need the distribution to integrate to 1!

Open Space Risk

Open Space Risk: the relative measure of open space to the full space

\[ R_o(f) = \frac{\int_{\mathcal{O}} f(x) \, dx}{\int_{S_o} f(x) \, dx} \]

Open space + positive training examples
The open set recognition problem

**Preliminaries**

Space of positive class data: \( \mathcal{P} \)

Space of other known class data: \( \mathcal{K} \)

Positive training data: \( \hat{\mathcal{V}} = \{v_1, \ldots, v_m\} \) from \( \mathcal{P} \)

Negative training data: \( \hat{\mathcal{K}} = \{k_1, \ldots, k_n\} \) from \( \mathcal{K} \)

Unknown negatives appearing in testing: \( \mathcal{U} \)

Testing data: \( \mathcal{T} = \{t_1, \ldots, t_z\}, t_i \in \mathcal{P} \cup \mathcal{K} \cup \mathcal{U} \)

Assume the problem openness is \( > 0 \)
The open set recognition problem

Minimize open set risk:

\[
\arg\min_{f \in \mathcal{H}} \left\{ R_\mathcal{O}(f) + \lambda_r R_\mathcal{E}(f(\hat{V} \cup \hat{K})) \right\}
\]

Open Space Risk Associated with \( \mathcal{U} \)

Regularization Constant

Empirical Risk Function
What options do we have to solve this problem?
Binary Classification
Multi-class 1-vs-All Classification
1-class Classification

Why didn’t the 1-class SVM catch on?

- Zhou and Huang *Multimedia Systems* 2003
  - Kernel and parameter selection
    - Gaussian kernels lead to over-fitting
    - Parameters chosen in *ad hoc* fashion
    - An issue in other domains too!

Other approaches


Let’s include open space risk in our optimization problem
Base Linear 1-vs-Set Machine
Generalization
Specialization
Open space risk for linear slab model

\[ \delta_A \text{ Marginal distance of near plane} \]

\[ \delta_\Omega \text{ Marginal distance of far plane} \]

\[ \delta^+ \text{ Separation needed to account for all positive data} \]

\[ \frac{\delta_\Omega - \delta_A}{\delta^+} \text{ Overgeneralization risk} \]

\[ \frac{\delta^+}{\delta_\Omega - \delta_A} \text{ Overspecialization risk} \]
Open space risk for linear slab model

\[ R_\varsigma = \frac{\delta_\Omega - \delta_A}{\delta^+} + \frac{\delta^+}{\delta_\Omega - \delta_A} + p_A \omega_A + p_\Omega \omega_\Omega \]

Two additional terms

Importance of open space around \( A \)
Importance of open space around \( \Omega \)

Margin around \( A \)
Margin around \( \Omega \)
Sketch of the 1-vs-Set Machine training algorithm

1. Train a linear SVM $f$ using $\hat{V}$ and $\hat{K}$

2. Generate decision scores for each training point in $\hat{V}$ and $\hat{K}$

3. Sort decision scores, where $s_k$ is the minimum and $s_j$ is the maximum

4. Initialize $A$ to margin plane of $f$, and $\Omega$ to $s_j$

5. Iteratively move $A$ to $s_{k+1}$ or $s_{k-1}$, $\Omega$ to $s_{j-1}$ or $s_{j+1}$ to minimize $R_\zeta(f) + \lambda_r R_\varepsilon$
1-vs-Set Machine Plane Refinement

Positive Pressure $p_A > 0$

Negative Pressure $p_A < 0$

Plane $A$ after initial optimization

Plane $A$ after refinement with $p_A = -0.5$
function PREDICT(t_x, f, A, \Omega) 
    if (A \leq f(t_x) \text{ and } f(t_x) \leq \Omega) then Return +1
    else Return -1
end if
end function
How can we evaluate open set recognition in a controlled manner?
Accuracy as a statistic for open set problems

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]

Imagine the following case:

1/100 $TP$ correct
100,000/100,000 $TN$ correct

99.9% accuracy!
F-measure as a statistic for open set problems

Consistent point of comparison across inconsistent precision and recall numbers:

\[
F\text{-measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]
Open Set Object Recognition

Cross-data set methodology*
Training: Caltech 256

Testing: Caltech 256 + ImageNet

Open Universe of 88 classes: 1 positive class, \( n \) training classes, 87 negative testing classes (532,400 images)

Open Universe of 212 classes: 1 positive class, \( n \) training classes, 211 negative testing classes (13,610,400 images)

Features

Histogram of Oriented Gradients


LBP-like descriptor

## 1-vs-Set Machine vs. Typical SVMs

<table>
<thead>
<tr>
<th>2-tailed paired t-test</th>
<th>binary 1-vs-Set</th>
<th>binary linear</th>
<th>binary RBF</th>
<th>1-class 1-vs-Set</th>
<th>1-class linear</th>
<th>1-class RBF</th>
</tr>
</thead>
<tbody>
<tr>
<td>binary 1-vs-Set (HOG 88)</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>binary linear (HOG 88)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>binary RBF (HOG 88)</td>
<td>—</td>
<td>++</td>
<td>—</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>1-class 1-vs-Set (HOG 88)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>**</td>
<td>—</td>
</tr>
<tr>
<td>1-class linear (HOG 88)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>1-class RBF (HOG 88)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>++</td>
<td>—</td>
</tr>
<tr>
<td>binary 1-vs-Set (HOG 212)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>1-class 1-vs-Set (HOG 212)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>*</td>
</tr>
<tr>
<td>binary 1-vs-Set (LBP-like 88)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>1-class 1-vs-Set (LBP-like 88)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>**</td>
<td>—</td>
</tr>
<tr>
<td>binary 1-vs-Set (LBP-like 212)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>1-class 1-vs-Set (LBP-like 212)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>**</td>
<td>—</td>
</tr>
</tbody>
</table>

** 1-vs-Set Machine is statistically significant at $p < 0.01$

* 1-vs-Set Machine is statistically significant at $p < 0.05$

++ Baseline Machine is statistically significant at $p < 0.01$

— No statistical significance
Top 25 classes for the open universe of 88 classes
Top 25 classes for the open universe of 88 classes
F-measure as a function of openness

![Graph showing F-measure as a function of openness for different models: Binary 1-vs-Set Machine, linear kernel, Binary SVM, linear kernel, and Binary SVM, RBF kernel. The graph displays the F-measure values for openness percentages ranging from 42% to 82%.](image)
Near and far plane pressures for open universe of 88 classes

The second plane has an impact on recognition performance.
Open Set Face Verification

Labeled Faces in the Wild

Genuine Pair

Impostor Pair

Impostor Pair

Impostor Pair

Gallery classes: 12 people with at least 50 images
Impostor classes: 82 other people in LFW
1,316 test images across all classes
Features: LBP-like and Gabor*

Open set face verification
Further Reading


Code

1-vs-Set Machine on GitHub:
https://github.com/tboult/libSVM-onevset

Data sets:
http://www.metarecognition.com/openset/