Optimal Asymmetrical SVM Using Pattern Search. A Health Care Application

G. Cohen¹ R. Meyer¹

¹Direction of Medico Economic Analysis
University Hospital of Geneva

G. Cohen
Oral presentation at MIE 2011,
August 30, 2011 Oslo
Outline

- Introduction
- Surveillance of nosocomial infections, methods
- Learning with imbalanced dataset
- SVM Classifier
- Hooke and Jeeves Pattern Search
- Experiments
- Results
- Conclusions and Future Work
**Nosocomial Infections (NIs)**

**NIs**: defined as those arising after 48 hours of hospitalization.

- Around 1.4 million patients per day are affected by NIs throughout the world (2009).
- In Europe, an estimated 5 million NIs at least occur in acute care hospitals annually, contributing to 135,000 deaths and representing around 25 million extra days of hospital stay (2009).

**Surveillance**: The tracking of NI is necessary to monitor overall effectiveness of Infection Control practices and health related patient and staff education.

**Methods**:

- prospective, ongoing surveillance (incidence studies)
- Trans-sectional assessment (i.e. prevalence studies)
Methodology & Datas

Methodology

Prevalence survey (800 hours for data collection and 100 hours for entering data in database)

Data preprocessing

- 688 patient records 83 variables $\rightarrow$ 683 patient records 49 variables
- Missing values replaced by
  - class-conditionnal mean (continuous variables)
  - or class-conditionnal mode (nominal variables)
- Distribution: 683 patients : 75 infected, 608 not infected.

Difficulty

Class imbalance (11%/89%).
Datasets used for training are not always balanced

Interesting or abnormal class is rare

Most of the real world problems are skewed: Medical diagnosis, Network intrusion detection, Direct marketing, Security: >99.99% of Americans are not terrorists . . . .

**How to handle Imbalance?**

- Sampling approaches (balance it!). Over/Down sampling the small/large class or combination of both,
- Cost sensitive classifier,
- SVMs approaches,
- Recognition-based learning or Novelty detection, ignore one of the two classes, learn from a single class

. . . .
Asymmetrical Non Linear Soft Margin SVM

\[
\begin{align*}
\min_{\xi \geq 0} & \quad \frac{1}{2} w^T w + C^+ \sum_{i \in C^+} \xi_i + C^- \sum_{i \in C^-} \xi_i \\
\text{s.t.} & \quad y_i (w^T x_i - b) + \xi_i \geq 1
\end{align*}
\]

\[
f(x) = w^T x + b = \sum_{i=1}^{n} \alpha_i \underbrace{\kappa(x_i, x)}_{\langle \phi(x_i), \phi(x) \rangle} + b
\]

- "Push the planes apart, and minimize weighted distance of misclassified points."
- Allows one to choose different C values for the two classes.
- Often used to weight rare class more heavily.
- Use "kernel function" (Trick: do not use \( \phi \), but \( \kappa \!): \kappa(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle \) ex: \( \kappa(x_i, x_j) = e^{-\|x_i - x_j\|^2 / 2\sigma^2} \)
Model selection for SVM

- Decision function: \( f(x) = w^T x + b = \sum_{i=1}^{n} \alpha_i \kappa(x_i, x) + b \)
- Goal:

\[
\text{Minimize: } R(f_\theta) = \int_{\mathcal{X} \times \mathcal{Y}} \ell(f_\theta(x), y) dP(x, y)
\]

- Bound or estimate \( R(f_\theta) \): Cost Weighted Accuracy
  \[ \text{CWA} = w \times \text{sensitivity} + (1-w) \times \text{specificity} \]
- Parameters \( \theta : (\theta_1, \ldots, \theta_n) \)
  - Scale of the kernel \( \gamma (= 1/2\sigma^2) \)
  - Regularization parameters \( C^+, C^- \)
- Traditionnal approaches
  - By hand
  - Grid Search
Method of Hooke and Jeeves (1961)

Considers only objective function: not its partial derivatives. Sequential technique each step of which consists of two kind of moves. One to explore the local behavior of the objective function, and the other to take advantage of the pattern direction.

**Figure:** Types of moves for Hooke and Jeeves search method: exploratory moves and pattern moves.
Method of Hooke and Jeeves

\[ f_1 = f(t_{k,j}) \]

Start \((j = 0)\)

\( j = j + 1 \)

Inc. coord. \( v = \Delta \theta_q \)
\( f_2 = f(t_{k,j-1} + v) \)

\( j > p \)

Dec. coord. \( v = -\Delta \theta_q \)
\( f_2 = f(t_{k,j-1} + v) \)

\( f_2 < f_1 \)

\( v = 0 \)
\( t_{k,j} = t_{k,j-1} + v \)

Start at \( b_k; t_{k0} = b_k; k = 1 \)

Exploratory move \((t_{k0})\)

Explanatory move \((t_{k0})\)

\( f(t_{kp}) < f(b_k) \)

\( \Delta \theta < \epsilon \)

Exit

Decrease step size
\( \gamma\{\Delta \theta\}; k = k+1; b_k = b_{k-1} \)

Set new base point
\( k = k + 1; b_k = t_{kp} \)

Pattern move
\( t_{k0} = b_k + \alpha(b_k - b_{k-1}) \)

G. Cohen, R. Meyer
Optimal Asymmetrical SVM Using Pattern Search
Figure: Illustration of the method of Hooke and Jeeves for the minization of \( f(x) = (x - 2)^4 + (x - 2y)^2 \). Starting point \((0, 3)\).
Experimental setup

Performance metrics
- Sensitivity, Specificity, Accuracy, Cost Weighted Accuracy ($w = 0.7$)

Evaluation strategy
- 5-fold stratified cross-validation
- RBF Gaussian kernel

$$\kappa(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}} = e^{-\gamma\|x_i - x_j\|^2}$$

Optimal values for hyperparameters: experiments with different classifiers using:
1. a grid: range of values: $C_{+/-} \in \{2^i\}_{i=-5,...,8}$ and $\gamma \in \{2^i\}_{i=-8,...,5}$
2. HJPS parameters: $\Delta \theta = 0.2$, $\gamma = 0.5$, $\epsilon = 10^{-5}$ and $\alpha = 1.2$
## Results

<table>
<thead>
<tr>
<th>Methods</th>
<th>CPU Time [min]</th>
<th>Nb. fonct. ev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>GS</td>
<td>36.17</td>
<td>2744</td>
</tr>
<tr>
<td>HJPS</td>
<td>1.35</td>
<td>13</td>
</tr>
</tbody>
</table>

**Figure**: Computational load: CPU time and number of function evaluations.

<table>
<thead>
<tr>
<th>Meth.</th>
<th>Parameters</th>
<th>Acc.</th>
<th>Sens.</th>
<th>Spec.</th>
<th>CWA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C^+$</td>
<td>$C^-$</td>
<td>$\sigma$</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>GS</td>
<td>0.5</td>
<td>0.06</td>
<td>$1.6 \times 10^{-2}$</td>
<td>83.59</td>
<td>81.33</td>
</tr>
<tr>
<td>HJPS</td>
<td>19.5</td>
<td>1.77</td>
<td>$12 \times 10^{-4}$</td>
<td>81.84</td>
<td>88</td>
</tr>
</tbody>
</table>

**Figure**: Performance of SVMs with optimal parameter set $(\gamma, C^+, C^-)$ found via GS and HJPS.
Conclusion

- The proposed algorithm can reliably find good ASVM models with RBF kernels in a fully automated way (robustness, simplicity and ease of implementation).
- Promising results: sensitivity 88%. We feel that it is a promising approach to the detection of nosocomial infections and can become a reliable component of an infection control system.

Future Work

- Plan to merge HJPS with Genetic Algorithms (GA) by first performing a coarse search for the global minimum by means of a GA and then refining the solution by a HJPS.
- prospectively validate the classification model obtained by performing in parallel a standard prevalence survey.