Applying One-vs-One and One-vs-All Classifiers in $k$-Nearest Neighbour Method and Support Vector Machines to an Otoneurological Multi-Class Problem

Kirsi Varpa, Henry Joutsijoki, Kati Iltanen, Martti Juhola
School of Information Sciences - Computer Science
University of Tampere, Finland
Introduction: From a Multi-Class Classifier to Several Two-Class Classifiers

- We studied how splitting of a multi-class classification task into several binary classification tasks affected predictive accuracy of machine learning methods.
  - One classifier holding nine disease class patterns was separated into multiple two-class classifiers.
- Multi-class classifier can be converted into
  - One-vs-One (OVO, 1-vs-1) or
  - One-vs-All the rest (OVA, 1-vs-All) classifiers.
From a Multi-Class Classifier to Several Two-Class Classifiers

nr of classifiers = 36 = \( nr \text{ of classes} \cdot (nr \text{ of classes} - 1) \)

nr of classifiers = 9 = nr of classes
One-vs-One (OVO) Classifier

- The results of each classifier are put together, thus having 36 class proposals (votes) for the class of the test sample.

- The final class for the test sample is chosen by the majority voting method, the max-wins rule: A class, which gains the most votes, is chosen as the final class.

  \[
  \begin{array}{cccccccccccccccccccccccccccc}
  2 & 3 & 3 & 4 & 5 & 6 & 7 & 8 & 1 & 2 & 5 & 6 & 7 & 8 & 9 & 1 & 5 & 3 & 7 & 5 & 6 & 1 & 2 & 4 & 8 & 5 & 1 & 7 & 3 & 4 \\
  1 & 8 & 9 & 1 & 2 & 1 \\
  \end{array}
  \rightarrow \text{max votes to class 1 (max-wins)}
  \]

  \[
  \begin{array}{cccccccccccccccccccccccccccc}
  2 & 3 & 3 & 4 & 5 & 6 & 7 & 8 & 1 & 2 & 5 & 6 & 7 & 8 & 9 & 1 & 5 & 3 & 7 & 5 & 6 & 1 & 2 & 4 & 8 & 5 & 6 & 7 & 3 & 4 \\
  1 & 8 & 9 & 1 & 2 & 9 \\
  \end{array}
  \rightarrow \text{max votes to classes 1 and 5 } \rightarrow \text{tie: SVM: 1-NN between tied classes 1 and 5,}

  \[
  \begin{array}{cccccccccccccccccccccccccccc}
  2 & 3 & 3 & 4 & 5 & 6 & 7 & 8 & 1 & 2 & 5 & 6 & 7 & 8 & 9 & 1 & 5 & 3 & 7 & 5 & 6 & 1 & 2 & 4 & 8 & 5 & 6 & 7 & 3 & 4 \\
  1 & 8 & 9 & 1 & 2 & 9 \\
  \end{array}
  \rightarrow \text{max votes to classes 1 and 5 } \rightarrow \text{tie: SVM: 1-NN between tied classes 1 and 5,}

  \[
  \begin{array}{cccccccccccccccccccccccccccc}
  k-\text{NN: nearest class (1 or 5) from classifiers 5-6, 1-3, 3-5, 1-4, 2-5, 5-8, 1-5, 5-9, 1-7 and 1-8.}
  \end{array}
  \]
One-vs-All (OVA) Classifier

- Each classifier is trained to separate one class from all the rest of the classes.
  - The class of the rest of the cases is marked to 0.
- The test sample is input to each classifier and the final class for the test sample is assigned according to the winner-takes-all rule from a classifier voting a class.

\[
\begin{bmatrix}
0 & 0 & 0 & 0 & 5 & 0 & 0 & 0 & 0
\end{bmatrix} \rightarrow \text{vote to a class 5 (winner-takes-all)}
\]

\[
\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix} \rightarrow \text{tie: find 1-NN from all the classes}
\]

\[
\begin{bmatrix}
0 & 2 & 0 & 0 & 6 & 0 & 0 & 0
\end{bmatrix} \rightarrow \text{votes to classes 2 and 6} \rightarrow \text{tie:}
\]

SVM: 1-NN between tied classes 2 and 6,

\[
k-\text{NN: nearest class (2 or 6) from classifiers 2-vs-All and 6-vs-All.}
\]
Data

- Classifiers were tested with an otoneurological data containing 1,030 vertigo cases from nine disease classes.
- The dataset consists of 94 attributes concerning a patient’s health status: occurring symptoms, medical history and clinical findings.
- The data had about 11 % missing values, which were imputed.

<table>
<thead>
<tr>
<th>Disease name</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acoustic Neurinoma</td>
<td>131</td>
<td>12.7</td>
</tr>
<tr>
<td>Benign Positional Vertigo</td>
<td>173</td>
<td>16.8</td>
</tr>
<tr>
<td>Meniere's Disease</td>
<td>350</td>
<td>34.0</td>
</tr>
<tr>
<td>Sudden Deafness</td>
<td>47</td>
<td>4.6</td>
</tr>
<tr>
<td>Traumatic Vertigo</td>
<td>73</td>
<td>7.1</td>
</tr>
<tr>
<td>Vestibular Neuritis</td>
<td>157</td>
<td>15.2</td>
</tr>
<tr>
<td>Benign Recurrent Vertigo</td>
<td>20</td>
<td>1.9</td>
</tr>
<tr>
<td>Vestibulopatia</td>
<td>55</td>
<td>5.3</td>
</tr>
<tr>
<td>Central Lesion</td>
<td>24</td>
<td>2.3</td>
</tr>
</tbody>
</table>
Methods

• OVO and OVA classifiers were tested using 10-fold cross-validation 10 times with
  • $k$-Nearest Neighbour ($k$-NN) method and
  • Support Vector Machines (SVM).
• Basic 5-NN method (using a classifier with all disease classes) was also run in order to have the baseline where to compare the effects of using multiple classifiers.
**k-Nearest Neighbour Method (k-NN)**

- *k*-NN method is a widely used, basic instance-based learning method that **searches for the *k*** most similar cases of a test case from the training data.
- In similarity calculation were used Heterogeneous Value Difference Metric (HVDM).
Support Vector Machine (SVM)

- The aim of SVM is to find a hyperplane that separates classes \( C_1 \) and \( C_2 \) and maximizes the margin, the distance between the hyperplane and the closest members of both classes.
- The points, which are the closest to the separating hyperplane, are called Support Vectors.
- Kernel functions were used with SVM because the data was linearly non-separable in the input space.
### Results

<table>
<thead>
<tr>
<th>Disease</th>
<th>Cases</th>
<th>5-NN %</th>
<th>5-NN %</th>
<th>SVM linear %</th>
<th>SVM RBF %</th>
<th>5-NN %</th>
<th>SVM linear %</th>
<th>SVM RBF %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acoustic Neurinoma</td>
<td>131</td>
<td>89.5</td>
<td>95.0</td>
<td>91.6</td>
<td>87.2</td>
<td>90.2</td>
<td>90.6</td>
<td>90.7</td>
</tr>
<tr>
<td>Benign Positional Vertigo</td>
<td>173</td>
<td>77.9</td>
<td>79.0</td>
<td>70.0</td>
<td>67.0</td>
<td>77.6</td>
<td>73.5</td>
<td>78.6</td>
</tr>
<tr>
<td>Meniere’s disease</td>
<td>350</td>
<td>92.4</td>
<td>93.1</td>
<td>83.8</td>
<td>90.1</td>
<td>89.8</td>
<td>87.8</td>
<td>91.5</td>
</tr>
<tr>
<td>Sudden Deafness</td>
<td>47</td>
<td>77.4</td>
<td>94.3</td>
<td>88.3</td>
<td>79.4</td>
<td>87.4</td>
<td>61.3</td>
<td>58.1</td>
</tr>
<tr>
<td>Traumatic vertigo</td>
<td>73</td>
<td>89.6</td>
<td>96.2</td>
<td>99.9</td>
<td>99.3</td>
<td>77.7</td>
<td>79.9</td>
<td>96.7</td>
</tr>
<tr>
<td>Vestibular Neuritis</td>
<td>157</td>
<td>87.7</td>
<td>88.2</td>
<td>82.4</td>
<td>81.4</td>
<td>85.0</td>
<td>85.4</td>
<td>84.3</td>
</tr>
<tr>
<td>Benign Recurrent Vertigo</td>
<td>20</td>
<td>3.0</td>
<td>4.0</td>
<td>20.0</td>
<td>16.5</td>
<td>8.0</td>
<td>21.0</td>
<td>8.0</td>
</tr>
<tr>
<td>Vestibulopatia</td>
<td>55</td>
<td>9.6</td>
<td>14.0</td>
<td>16.5</td>
<td>22.8</td>
<td>15.8</td>
<td>15.3</td>
<td>13.5</td>
</tr>
<tr>
<td>Central Lesion</td>
<td>24</td>
<td>5.0</td>
<td>2.1</td>
<td>26.0</td>
<td>28.5</td>
<td>15.0</td>
<td>19.0</td>
<td>15.8</td>
</tr>
<tr>
<td>Median of True Positive Rate (%)</td>
<td>77.9</td>
<td>88.2</td>
<td>82.4</td>
<td>79.4</td>
<td>77.7</td>
<td>73.5</td>
<td>78.6</td>
<td></td>
</tr>
<tr>
<td>Total Classification accuracy (%)</td>
<td>79.8</td>
<td>82.4</td>
<td>77.4</td>
<td>78.2</td>
<td>78.8</td>
<td>76.8</td>
<td>79.4</td>
<td></td>
</tr>
</tbody>
</table>

Linear kernel with box constraint $bc = 0.20$ (OVO and OVA)
Radial Basis Function (RBF) kernel with $bc = 0.4$ and scaling factor $\sigma = 8.20$ (OVO), $bc = 1.4$ and $\sigma = 10.0$ (OVA)
Conclusions

• The results show that in most of the disease classes the use of multiple binary classifiers improves the true positive rates of disease classes.
• The results show that in most of the disease classes especially, 5-NN with OVO classifiers worked out better with this data than 5-NN with OVA classifiers.
Thank you for your attention!

Questions?

Kirsi.Varpa@cs.uta.fi

More information about the subject:

Allwein EL, Schapire RE, Singer Y. Reducing multiclass to binary: a unifying approach for margin