Efficient pairwise classification using Local Cross Off strategy

Mohammad ali Bagheri
Binary vs. Multiclass Classification

- Real word applications

- Class binarization
  - One-versus-all (OVA)
  - One-versus-one (OVO)
  - Error Correcting Output Codes (ECOC)
Error Correcting Output Codes

- Idea: designing a codeword for each of the classes
- matrix $M$ of size $L \times Nc$: each cell is $\{-1, +1\}$
  - Column ---> dichotomy classifier
  - Row: is a unique codeword that is associated with an individual target class

- Sparse ECOC
  - Adding 0 to the matrix

<table>
<thead>
<tr>
<th>Class</th>
<th>$h_1$</th>
<th>$h_2$</th>
<th>$h_3$</th>
<th>$h_4$</th>
<th>$h_5$</th>
<th>$h_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>$\omega_4$</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Drawbacks of OVO

- **incompetent classifiers**

Suppose a problem with 4 classes

- new test instance belongs to C3
- Training phase: 1vs2, 1vs3, 1vs4, 2vs3, 2vs4, 3vs4
- Testing phase:
  - $h_{12} \rightarrow 1$
  - $h_{13} \rightarrow 3$
  - $h_{14} \rightarrow 1$
  - $h_{23} \rightarrow 2$
  - $h_{24} \rightarrow 4$
  - $h_{34} \rightarrow 3$

Several methods has been proposed: A&O, CC, …
Proposed Method

- **Training phase**: build pair classifiers

- **Test phase**:
  - Define local neighborhood
  - Figures out which classes are the most frequent in those neighbors
  - Choose relevant classifiers based on the class frequency
Main idea: remove the irrelevant classifiers

*Local Cross Off*

- **LCO-Version 1:**
  - The two most frequent classes of the nearest $K$ neighbors in the training set of each test pattern are found.
  - One binary classifier is selected to classify test pattern.

- **LCO-Version 2:**
  - All target classes of the nearest $K$ neighbors in the training set of each test pattern are found.
  - Classifiers that correspond to all pairwise combinations of these classes are then nominated.
  - Majority voting.
Validation over benchmark datasets

- **Methods:**
  - OVO, OVA, A&O, and ECOC

- In modified - nearest neighbor algorithm: \( K = 5 \)

- **Base learners:**
  - Linear Support Vector Machine
  - Multilayer Perceptron (MLP).

- **Evaluation**
  - Accuracy based on 10-fold cross-validation
  - fair comparison!
Validation over benchmark datasets

Pair accuracy comparison:

- OVO
- OVA
- A&O
- dense ECOC
- Sparse ECOC

[Graph showing accuracy comparison for various datasets]
Validation over benchmark datasets: Statistical analysis

- Recommendations of Demsar: non-parametric tests

- General procedure:
  - Iman–Davenport test --- Nemenyi test

- Iman–Davenport test:
  - rank competing methods for each dataset
  - The method’s mean rank by averaging its ranks across all experiments
  - Applying the Iman–Davenport formula
Validation over benchmark datasets

- Nemenyi test - SVM

![Graph showing validation results for different datasets and methods. The x-axis represents different datasets, and the y-axis shows the validation scores. The methods include LCO_v2, sparse ECOC, dense ECOC, A&O, 1vsAll, and 1vs1.]
Validation over benchmark datasets

- Nemenyi test - MLP

Graph showing the performance comparison of different methods including LCO_v2, sparse ECOC, dense ECOC, A&O, 1vsAll, and 1vs1.
Conclusions

- We presented a novel strategy for pairwise classification approach to deal with multiclass problems.

- The proposed technique is based on omitting the votes of irrelevant binary classifiers, in order to improve final classification accuracy.

- The proposed LCO method validated over a set of benchmark dataset.
Conclusions

- The experimental evaluation shows some strong and consistent evidence of performance improvements compared to the one-versus-one, one-versus-all, A&O, and ECOC methods.

- The main reason behind this improvement is that the LCO approach is benefited from efficient nearest neighbor rule as a preprocessing step in pairwise structure and the strength of the other adapted powerful binary classifiers.
Thanks
Questions