Intelligent Feature Extraction for Ensemble of Classifiers

Paulo V. W. Radtke
Robert Sabourin
Tony Wong

Summary

- Motivation
- Problem Statement
- IFE Methodology
- EoC Optimization
- Experimental Results
- Conclusions

Motivation

- An Ensemble of Classifiers (EoC) can provide higher accuracy than a single classifier
- EoCs are usually created running specific learning algorithm (boosting, bagging, random subspace method, etc)

Problem Statement

- Key issue: to generate and select a set of diverse and fairly accurate base classifiers
- Two-level approach:
  - Intelligent Feature Extraction (IFE)
  - Classifier Selection (EoC)
- Multi-objective Genetic Algorithms
- Experiments on isolated handwritten digits recognition
System overview

- IFE
- MOMA

Representation
Set RS

Selection
[ELO'2005]

EoC Optimization
[ICDAR'2005]

Single best representation

Experimental Protocol

- Test the IFE on a known problem:
  - Isolated handwritten digits
  - 22 features extracted per zone [Oliveira al., PAMI’02]
  - IFE: Zoning & feature extraction (without FSS)

- IFE & EoCs processes implemented with MOMA

Experimental Protocol

- MOMA
  - One point crossover operator (Pc = 0.8)
  - Bitwise mutation operator (Pm = 1/L)
  - Local search:
    - NI = 3 iterations
    - Neighborhood size = 1
  - Archive size:
    - 5 solutions/slot
  - Population size:
    - IFE = 64
    - EoC = 186
  - Maximum search time:
    - 1 000 generations

Experimental Protocol

- Digits extracted from the NIST SD19 database:
  - learn/valid and validation: used to train the PD and MLP classifiers
  - optimization: used to evaluate classifier accuracy during optimization
  - selection: post-processing of candidate solutions
  - test1: hsf_4 [not showed in the icdar2005 proceedings]
  - test2: hsf_7
**Experimental Protocol**

1. Optimize IFE using MOMA
   - PD classifier in a wrapper mode (learn/validation databases)
   - Fitness evaluated with the optimization database

2. Select a single best solution SB from RS
   - Train |RS| = p classifiers (learn/validation databases)
   - Select the best solution with the selection database

3. Optimize a PD EoC and MLP EoC using MOMA
   - |K| = p base classifiers created from RS (learn/validation databases)
   - Fitness evaluated with the optimization database
   - Selection of the best EoC in K' with the selection database

**Experimental Results**

**Baseline representation**

- Single best representation found by the IFE & PD classifier

**Note1:** IFE found |RS| = 82 solutions

**Note2:** The same representation was also found with the MLP classifier

**Experimental Results**

- **PD classifier results:**
  - Classifier: Baseline, SB_PD, EoC_PD
  - |G|: 132, 330, -
  - $e_{avg}$: 3.01%, 2.31%, 1.51%
  - $E_{test\ hsf\ 4}$: 6.83%, 5.47%, 5.06%
  - $E_{test\ hsf\ 7}$: 2.96%, 2.18%, 1.94%
  - |EoC|: -

**Experimental results not showed in the ICDAR2005 proceedings**

- **MLP classifier results:**
  - Classifier: Baseline, SB_MLP, EoC_MLP
  - |G|: 132, 330, -
  - $e_{avg}$: 0.44%, 0.41%, 0.37%
  - $E_{test\ hsf\ 4}$: 2.89%, 2.44%, 2.37%
  - $E_{test\ hsf\ 7}$: 0.89%, 0.79%, 0.74%
  - |EoC|: -

**Experimental results not showed in the ICDAR2005 proceedings**
Computational Burden

- Beowulf cluster, 25 machines (Athlon XP 2300+), Linux Red Hat 7.3, MPI v7.
- IFE Optimization: 36 hours (estimated)
  - Synchronous Master-Slave, without load balancing
  - 64,000 solutions (PD classifiers)
- MLP training (|K|=82 classifiers): 3 weeks.
  - Asynchronous Master-Slave, with load balancing
- EoC optimization (PD/MLP): 30 minutes

Conclusions

- The IFE methodology outperformed the traditional human expert approach on a known problem
- Expert knowledge on the PR domain can be taken into account more easily in the feature extraction process
- The solution set RS (representations) produced by IFE is effective for EoC creation

Conclusions

- MOMA proved itself suitable for the IFE methodology and EoC creation
- MOMA performs better than traditional multi-objective Pareto-based optimization algorithms when applied in machine learning
- To avoid over-fitting, a selection (validation) database to select solutions in the archive IS MANDATORY

Questions?
Databases

<table>
<thead>
<tr>
<th>Database</th>
<th>Origin</th>
<th>Size</th>
<th>Subset definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>learn</td>
<td>hsf_0123</td>
<td>150 000</td>
<td>1- 150 000</td>
</tr>
<tr>
<td>learn'</td>
<td>hsf_0123</td>
<td>50 000</td>
<td>1- 50 000</td>
</tr>
<tr>
<td>validation</td>
<td>hsf_0123</td>
<td>15 000</td>
<td>150 001 – 165 000</td>
</tr>
<tr>
<td>optimization</td>
<td>hsf_0123</td>
<td>15 000</td>
<td>165 001 – 180 000</td>
</tr>
<tr>
<td>selection</td>
<td>hsf_0123</td>
<td>15 000</td>
<td>180 001 – 195 000</td>
</tr>
<tr>
<td>test_1</td>
<td>hsf_7</td>
<td>60 089</td>
<td></td>
</tr>
<tr>
<td>test_2</td>
<td>hsf_4</td>
<td>58 646</td>
<td></td>
</tr>
</tbody>
</table>

IFE Motivation

- Need to adapt pattern recognition (PR) systems to different problems related to the same domain
  - Ex: Isolated Numerals Rec. vs Upper Case Letters Rec.
- Lack of methodologies to extract and adapt automatically representations to several PR problems who belong to the same domain
- Intelligent Feature Extraction (IFE) methodology

IFE Methodology

- Case Study: Handwritten symbols modeled with features extracted from foci of attention in images
- Three hierarchical operations: zoning, feature extraction and feature subset selection
- Human knowledge is related to:
  - the zoning mechanism (location & size and number of retinas)
  - the choice of feature extraction operators

IFE Methodology

Feature extraction operators:
- Concavities (13 features)
- Contour (8 features)
- Surface (1 feature)

Feature subset selection operator:
- 22 bits binary string
- 1: active feature
- 0: inactive feature

Dividers Zoning

<table>
<thead>
<tr>
<th>0</th>
<th>0</th>
<th>1</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
IFE Methodology

- Uses a wrapper approach:
  - Projection Distance classifier (PD, Kimura et al., 1998)

- Hypothesis: representations found with PD classifier can be used to train more discriminant classifiers

- Optimization objectives:
  - Minimize classifier error rate
  - Minimize the feature set cardinality (e.g. number of zones)

Solution Selection

- IFE:
  - Provides the N best solutions for each cardinality in the Archive

- A post-processing stage selects the best solution based on the classifier’s capacity to generalize well on unseen data

- This step is performed to remove the over-fitting in the optimization stage

MOMA

- Multi-Objective Memetic Algorithm (MOMA)
  - Memetic Algorithm: GA & Local search mechanism

- MOMA is well suited for the IFE methodology:
  - Modified genetic selection strategy
  - Archiving strategy is considered during the optimization process
  - Local search mechanism (RRT algorithm) to improve convergence

- Two-objective optimization:
  - $o_1$: classifier error rate
  - $o_2$: feature set cardinality

MOMA

- Genetic selection emphasizes the best solutions for each $o_1$ value, regardless of domination.

Non dominated solutions
The IFE generates the set $RS = \{G_1, \ldots, G_p\}$.

We use $RS$ to create the set $K = \{K_1, \ldots, K_p\}$ of base classifiers.

To encode the EoC, we use a binary string $E$ of $p$ bits.
- The $i$th bit in $E$ indicates if classifier $K_i$ is selected (1) or not (0).

At the end of the optimization process, MOMA produces a set $K'$ of EoCs.

Two objectives guide the optimization:
- $o_1$: EoC cardinality
- $o_2$: EoC combined error rate

Fusion function:
- PD EoC: simple majority voting
- MLP EoC: output averaging

PD EoC – Selected Solutions: 2, 2, 3, 4, 4, 4, 5, 6, 6, 8, 8, 12, 12, 15, 15, 15, 15, 18, 20, 24 (Nb of zones)
PD EoC – Selected Solutions

7.66%  13.27%  7.53%  5.02%  3.53%  3.80%  3.27%
2.88%  2.62%  3.06%  2.59%  2.72%  2.57%  2.43%
2.46%  2.18%  2.36%  2.37%  2.35%
(error rates on hsf_7)

MLP EoC – Selected Solutions

8  10  18  30
Representations selected for each base classifier
(Nb of zones)

MLP EoC – Selected Solutions

0.9%  0.83%  0.86%  0.87%
Individual performance of base classifiers
(error rates on hsf_7)