LIVIA
Laboratoire d'imagerie, de vision et d'intelligence artificielle

Intelligent Feature Extraction for Ensemble of Classifiers

Paulo V. W. Radtke
Robert Sabourin
Tony Wong

Université du Québec
École de technologie supérieure

Summary

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

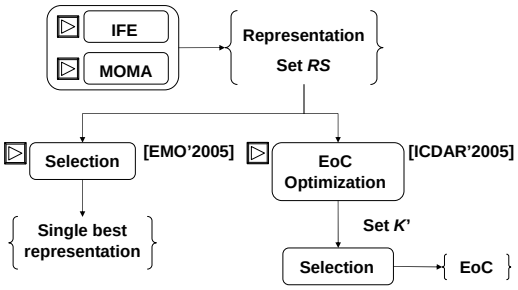
Motivation

- 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



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

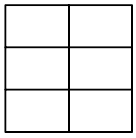
- Digits extracted from the NIST SD19 database:
  - *learn'/learn* and *validation*: used to train the PD and MLP classifiers
  - *optimization*: used to evaluate classifier accuracy during optimization
  - *selection*: post-processing of candidate solutions
  - $test_1$: hsf_4 [not showed in the icdar2005 proceedings]
  - $test_2$: hsf_7

Experimental Protocol

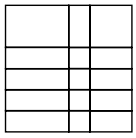
1. Optimize IFE using MOMA
   - PD classifier in a wrapper mode (learn'/validation databases)
   - Fitness evaluated with the optimization database
2. Select a single best solution SB from RS
   - Train $|RS| = p$ classifiers (learn/validation databases)
   - Select the best solution with the selection 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 (human expert)



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	$\|G\|$	e_{sel}	$E_{test\ hsf\_4}$	$E_{test\ hsf\_7}$	$\|EoC\|$
Baseline	132	3.01%	6.83%	2.96%	-
SB_{PD}	330	2.31%	5.47%	2.18%	-
EoC_{PD}	-	1.51%	5.06%	1.94%	19

Experimental results not showed in the ICDAR2005 proceedings

Experimental Results

- MLP classifier results:

Classifier	$\|G\|$	e_{sel}	$E_{test\ hsf\_4}$	$E_{test\ hsf\_7}$	$\|EoC\|$
Baseline	132	0.44%	2.89%	0.89%	-
SB_{MLP}	330	0.41%	2.44%	0.79%	-
EoC_{MLP}	-	0.37%	2.37%	0.74%	4

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

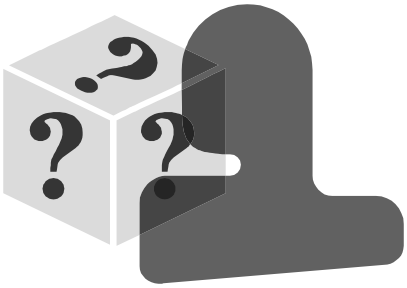
Conclusions

- The IFE methodology outperformed the traditional human expert approach on a known problem
- Expert knowledge on the PR domain can be take 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

Database	Origin	Size	Subset definition
learn	hsf_0123	150 000	1- 150 000
learn'	hsf_0123	50 000	1- 50 000
validation	hsf_0123	15 000	150 001 – 165 000
optimization	hsf_0123	15 000	165 001 – 180 000
selection	hsf_0123	15 000	180 001 – 195 000
test$_1$	hsf_7	60 089	
test$_2$	hsf_4	58 646	

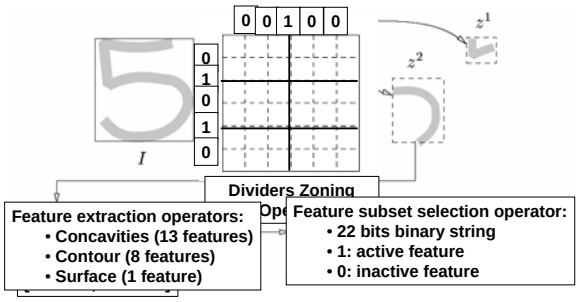
IFE Motivation

- Need to adapt pattern recognition (PR) systems to different problems related to the same domain
  - Ex: Isolated Numerals Rec. ⇨ 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



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)

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

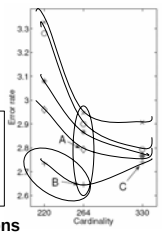
Solution Selection

- 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

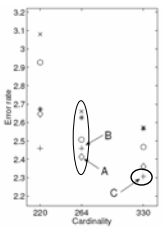
MOMA

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

Optimization

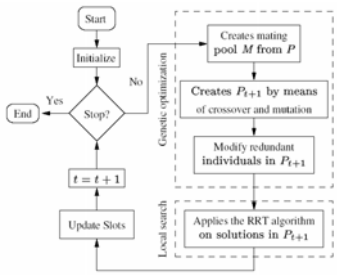


Unseen data



Non dominated solutions

MOMA



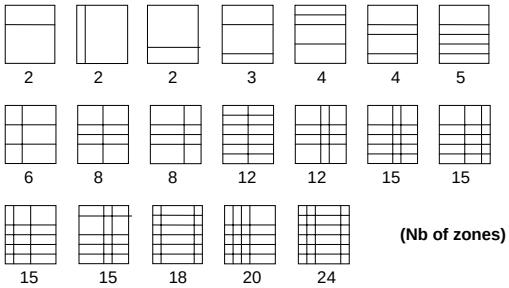
EoC optimization: MOMA

- The IFE generates the set *RS={G¹,..., Gᵖ}*
- We use RS to create the set *K={K¹,..., Kᵖ} of base classifiers*
- To encode the EoC, we use a binary string *E* of p bits
  - The ith 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

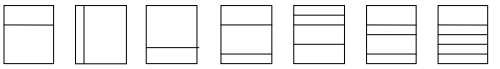
EoC optimization: MOMA

- 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

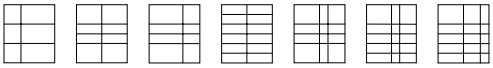
PD EoC – Selected Solutions



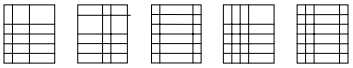
PD EoC – Selected Solutions



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%

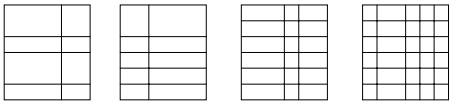
(error rates on hsf_7)

30

LIVIA - École de Technologie Supérieure

MLP EoC – Selected Solutions

Representations selected for each base classifier
(Nb of zones)



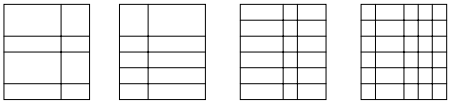
8 10 18 30

31

LIVIA - École de Technologie Supérieure

MLP EoC – Selected Solutions

Individual performance of base classifiers
(error rates on hsf_7)



0.9% 0.83% 0.86% 0.87%

32

LIVIA - École de Technologie Supérieure