Modulated Training of Cascaded Ensembles

by

Teo Susnjak
Massey University
New Zealand
Outline

- Introduction
- Problem outline
- Machine learning with PSL
- Training with positive sample bootstrapping
- Experimental results and limitations
- Adaptive learning with PSL-like frameworks
- Concluding remarks
- References
Ensemble Based Learning

- Trains multiple classifiers
- Combines collective decisions into a single classification
- Underlying principle: combined decision better than that of any individual expert

Key components:
  - Ensemble diversity
  - Methods to combine decisions
Cascades of Boosted Ensembles

- Weak classifiers arranged into layers
- Each layer classifies samples of increasing complexity
- Enables real-time classification
- Facilitates training with large negative datasets
- Modularization and tractability
Cascades of Boosted Ensembles

• Disadvantages
  - Training speed
  - Slow convergence to layer targets
  - Positive dataset size limitation
  - Cascade optimization issues
**PSL Training Structure**

- *Parallel strong* classifiers within the same *layer*
- Quasi dual-cascaded structure
- Faster convergence to layer targets
- Overall training runtime reduction
- Simplified optimization
PSL Training Structure

Propagation of Positive Samples at Training

Input 100% positive samples

Layer 1

% of correctly predicted
60%

% of incorrectly predicted
40%

20%

100% positive prediction for the layer

Propagation of Negative Samples at Training

Input 100% negative samples

% of correctly predicted
25%

% of incorrectly predicted
12.5%

100% negative propagation

combined correct prediction of 50%

combined 50% false acceptance rate for the layer
Bootstrapping Positives

Bootstrapping Procedure in the BDC Framework

Inter-cascade layer n - 1

Inter-cascade layer n

Inter-cascade layer n + 1

base set

base set

base set

Intra-layer stage 1

Intra-layer stage 2

Intra-layer stage 3

fill with new positives

apply stage 1 classifier

negative prediction

positive prediction

reserve set
Bootstrapping Positives - Results

Training Runtimes

CPU Seconds (accuracy to + 1%)

1e+07
8e+06
6e+06
4e+06
2e+06
0

BDC (max stage size 5)
BDC (max stage size 10)
PSL (max stage size 5)
PSL (max stage size 10)
Viola-Jones Cascaded

Classifier Type

5000/2000 Dataset
10000/2000 Dataset
15000/2000 Dataset
Bootstrapping Positives - Results

Rate of Weak Classifier Generation

Number of weak classifiers vs. Days

Classifiers:
- BDC
- PSL
- Viola-Jones Cascaded
Bootstrapping Positives - Overfitting

Example of samples trained in initial intra-layer stages

Samples trained in trailing intra-layer stages
Bootstrapping Positives - Overfitting

Inter-cascade layer n

- Base set
  - Negative prediction
    - Intra-layer stage 1
    - Intra-layer stage 2
    - Intra-layer stage 3
    - Intra-layer stage 4
Bootstrapping Positives - Results

ROC Graph Comparing BDC Classifiers with Adjustments for Overfitting for datasets of 10000

Hit Rate vs. False Detection Rate

Classifiers:
- PSL 10
- BDC_10
- BDC_10 with anti overfit adjustment
- BDC_10 with 2000 sized base sets
- Viola-Jones Cascaded
Adaptive Learning with PSL

- Statically trained classifiers insufficient
- Concept drift
  - Unpredictable changes in the underlying distribution of data
  - Abrupt or gradual
- Adaptability requirements:
  - Timely
  - No access to prior datasets
  - Balance between plasticity and stability
Adaptive Learning with PSL

Concept Learning Framework

Layer n - 1

Candidate sample

Layer n

Training Dataset (static)

Test Dataset (non-stationary)

$\alpha$ layer n

$\alpha^1 + \alpha^2 + \alpha^3 + \alpha^4 < \alpha^{layer n}$

then reject sample

Layer n + 1
Conclusion

• Cascades of ensembles can be modularized
• Modularization introduces tractability to training
• The new framework enables bootstrapping of positive samples
• Scope for extending for effective concept-drift handling and incremental learning
References (partial)


Questions?
Thank you!