# Dynamic Classifier Systems for Classifier Aggregation

David Štefka

Seminář z umělé inteligence
11. 12. 2008

Martin Holeňa

# Outline

# Classification

- classification – process of assigning *patterns* into *classes*
- $\mathcal{X}$ – feature space, $\vec{x}$ pattern
- $C_1, \ldots, C_N$ – classes
- classifier – mapping $\phi : \mathcal{X} \rightarrow [0,1]^N$
  $\phi(\vec{x}) = (\mu_1(\vec{x}), \ldots, \mu_N(\vec{x}))$
- $\mu_i(\vec{x})$ – degree of classification $x \in C_i$
- interpretation of $\mu_i(\vec{x})$ – depends on the classifier used (probability, fuzzy membership, ...)

# Classifier Combining

- method for improving the classification by using multiple classifiers and combining their outputs
- create a team $\mathcal{T} = (\phi_1, \ldots, \phi_r)$ of classifiers (bagging, boosting, . . . )
- aggregate the team using aggregator $\mathcal{A}$
- most of the aggregation methods are *static*

$$\vec{x} \begin{array}{l} \nearrow \phi_1 \\ \rightarrow \phi_2 \\ \quad \vdots \\ \searrow \phi_r \end{array} \longrightarrow \mathcal{T}(\vec{x}) \longrightarrow \mathcal{A} \longrightarrow \Phi(\vec{x})$$

# Classification example

- $C_1$ – bananas, $C_2$ – apples

# Classification example ctnd.

- $\phi(\vec{x}) = (0.1, 0.8)$

# Classification example ctnd.

- $\phi(\vec{x}) = (0.9, 0.1)$

# Classification example ctnd.

- $\phi(\vec{x}) = (0.1, 0.2)$
- Low *confidence*

# Classification Confidence

- degree of trust we can give to the classifier $\phi$
- probability of correct classification of $\vec{x}$ by $\phi$
- ability to answer "I don't know"
- confidence measure $\kappa_\phi : \mathcal{X} \longrightarrow [0, 1]$
- static measures – constant of the classifier (e.g., accuracy)
- dynamic measures – adapted to the currently classified pattern (e.g., local accuracy)

# Global Accuracy - GA

- static confidence measure
- validation (training) set $\mathcal{M}$
- proportion of patterns $\vec{y} \in \mathcal{M}$ correctly classified by $\phi$

$$\kappa_\phi^{(GA)}(\vec{x}) = \frac{\sum_{\vec{y} \in \mathcal{M}} I(\phi_{cr}(\vec{y}) \stackrel{?}{=} c(\vec{y}))}{|\mathcal{M}|}$$

# Euclidean Local Accuracy - ELA

- dynamic confidence measure
- validation (training) set $\mathcal{M}$
- $N(x)$ patterns from $\mathcal{M}$ neighboring with $\vec{x}$ (e.g., 20 nearest under Euclidean metric)
- proportion of patterns $\vec{y} \in N(\vec{x})$ correctly classified by $\phi$

$$\kappa_{\phi}^{(ELA)}(\vec{x}) = \frac{\sum_{\vec{y} \in N(\vec{x})} I(\phi_{cr}(\vec{y}) \stackrel{?}{=} c(\vec{y}))}{|N(\vec{x})|}$$

# Euclidean Local Match - ELM

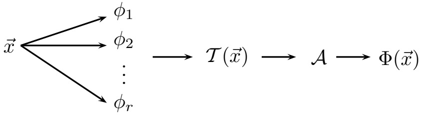
- dynamic confidence measure
- validation (training) set $\mathcal{M}$
- $N(x)$ patterns from $\mathcal{M}$ neighboring with $\vec{x}$ (e.g., 20 nearest under Euclidean metric)
- proportion of patterns $\vec{y} \in N(\vec{x})$ from the same class as $\phi$ is predicting for $\vec{x}$

$$\kappa_{\phi}^{(ELM)}(\vec{x}) = \frac{\sum_{\vec{y} \in N(\vec{x})} I(\phi_{cr}(\vec{x}) \stackrel{?}{=} c(\vec{y}))}{|N(\vec{x})|}$$

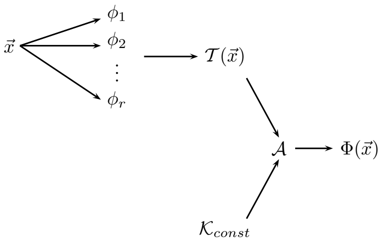
# Classifier Systems

- $\mathcal{S} = (\mathcal{T}, \mathcal{K}, \mathcal{A})$ – classifier system
- $\mathcal{T} = (\phi_1, \ldots, \phi_r)$ – classifiers
- $\mathcal{K} = (\kappa_{\phi_1}, \ldots, \kappa_{\phi_r})$ – confidence measures
- $\mathcal{A}$ – aggregator
- 3 types of classifier systems
  - confidence-free
  - static
  - dynamic

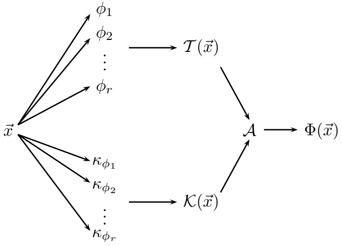
# Types of classifier systems




(a) Confidence-free




(b) Static




(c) Dynamic

# Mean value based aggregators

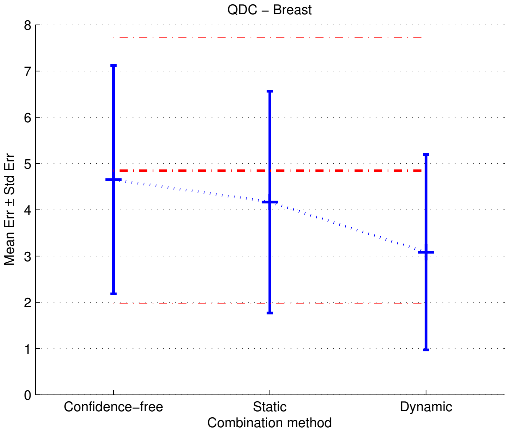
- confidence-free – mean value of the classifier outputs
- static – weighted mean; weights are static confidences
- dynamic – weighted mean; weights are dynamic confidences

$$\mathcal{T}(\vec{x}) = \begin{pmatrix} \phi_1(\vec{x}) \\ \phi_2(\vec{x}) \\ \vdots \\ \phi_r(\vec{x}) \end{pmatrix} = \begin{pmatrix} \mu_{1,1}(\vec{x}) & \mu_{1,2}(\vec{x}) & \dots & \mu_{1,N}(\vec{x}) \\ \mu_{2,1}(\vec{x}) & \mu_{2,2}(\vec{x}) & \dots & \mu_{2,N}(\vec{x}) \\ & & \ddots & \\ \mu_{r,1}(\vec{x}) & \mu_{r,2}(\vec{x}) & \dots & \mu_{r,N}(\vec{x}) \end{pmatrix}, \ \mathcal{K}(\vec{x}) = \begin{pmatrix} \kappa_{\phi_1}(\vec{x}) \\ \kappa_{\phi_2}(\vec{x}) \\ \vdots \\ \kappa_{\phi_r}(\vec{x}) \end{pmatrix}$$

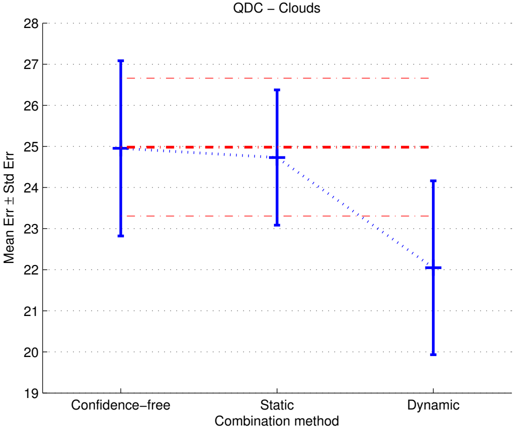
# Experiments

- comparison: confidence-free vs. static vs. dynamic systems
- systems of Quadratic Discriminant Classifiers
- various confidence measures
- aggregators based on mean value aggregator
- 4 artificial, 4 real-world datasets
- 10-fold crossvalidation
- ELM confidence measure

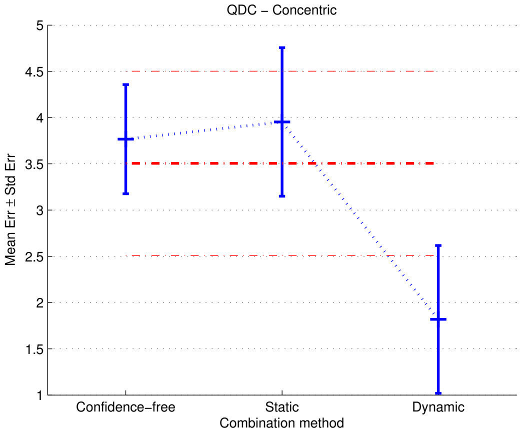
# Results – QDC


QDC – Breast
Mean Err ± Std Err
8
7
6
5
4
3
2
1
0
Confidence–free
Static
Dynamic
Combination method

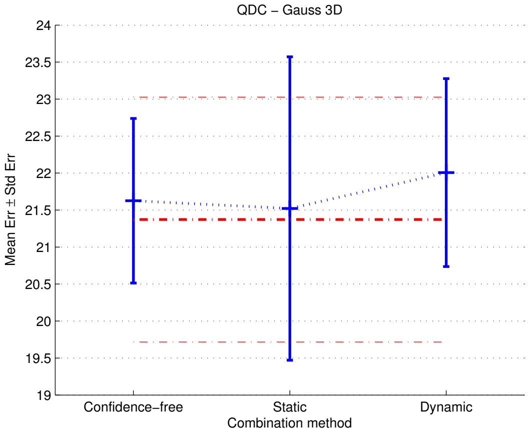
# Results – QDC


QDC – Clouds
Mean Err ± Std Err
28
27
26
25
24
23
22
21
20
19
Confidence−free
Static
Dynamic
Combination method

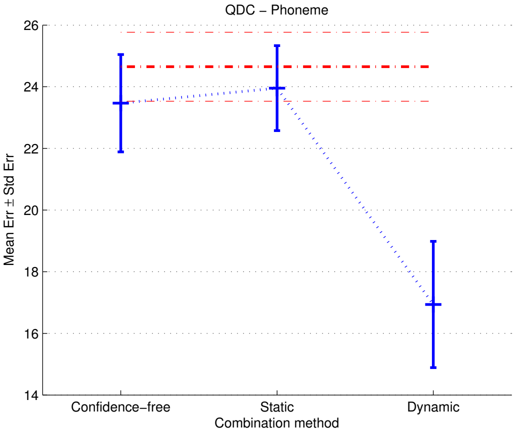
# Results – QDC


QDC – Concentric
Mean Err ± Std Err
5
4.5
4
3.5
3
2.5
2
1.5
1
Confidence–free
Static
Dynamic
Combination method

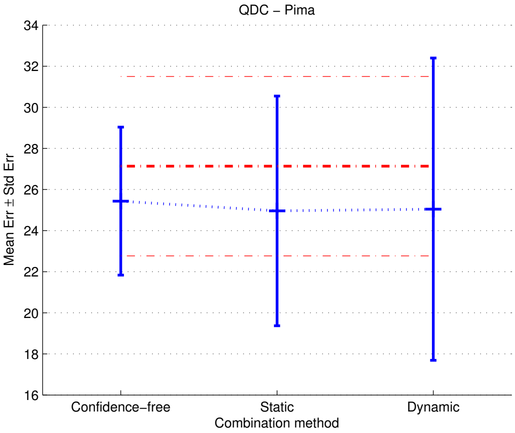
# Results – QDC


QDC – Gauss 3D
Mean Err ± Std Err
24
23.5
23
22.5
22
21.5
21
20.5
20
19.5
19
Confidence–free
Static
Dynamic
Combination method

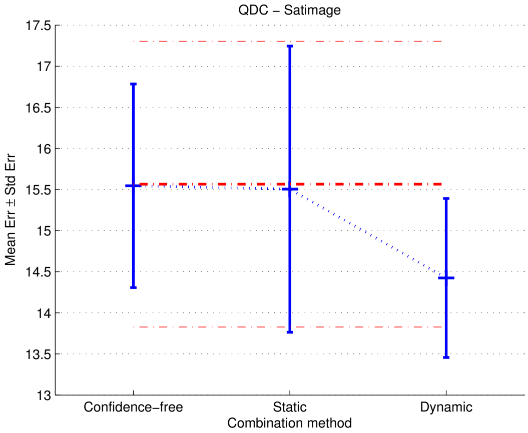
# Results – QDC


QDC – Phoneme
Mean Err ± Std Err
26
24
22
20
18
16
14
Confidence–free
Static
Dynamic
Combination method

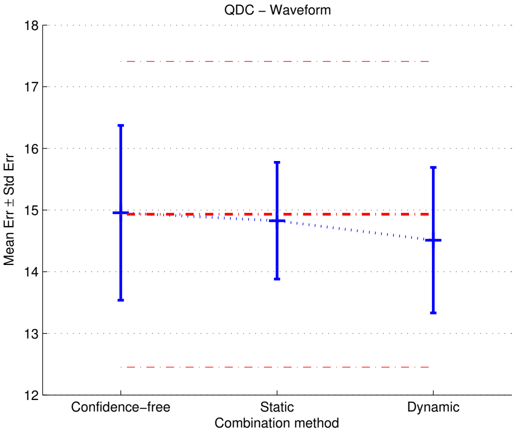
# Results – QDC


QDC – Pima
Mean Err ± Std Err
34
32
30
28
26
24
22
20
18
16
Confidence–free
Static
Dynamic
Combination method

# Results – QDC


QDC – Satimage
17.5
17
16.5
16
15.5
15
14.5
14
13.5
13
Mean Err ± Std Err
Confidence–free
Static
Dynamic
Combination method

# Results – QDC


QDC – Waveform
18
17
16
15
14
13
12
Mean Err ± Std Err
Confidence–free
Static
Dynamic
Combination method

# Experiments - Results

- experiments with QDC ensembles: dynamic classifier systems can significantly improve classification quality
- similar results with Random Forests on 17 datasets
- improvement in classification quality vs. higher computational complexity
- ELM more successfull than ELA

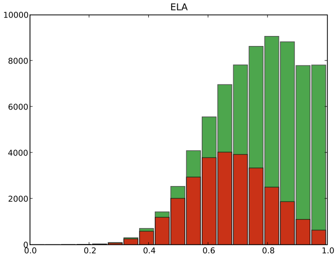
# Assessing Confidence Measures

- which confidence measure is better?
- will a confidence measure bring improvement in the classification quality?
- are the benefits of a dynamic classifier system worth bigger compuational complexity?

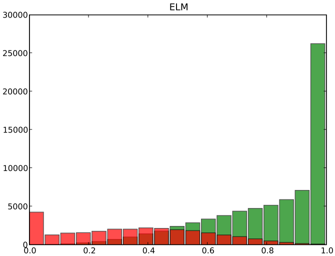
# OK/NOK Histograms

- confidence measure = probability of correct classification?
- distribution of correctly classified patterns (OK)
- distribution of correctly classified patterns (NOK)
- OK and NOK should not overlap and should be separated

# OK/NOK Histograms – Waveform

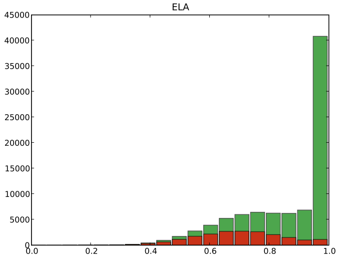



(d) ELA – bad separation




(e) ELM – relatively good separation

# OK/NOK Histograms – Phoneme

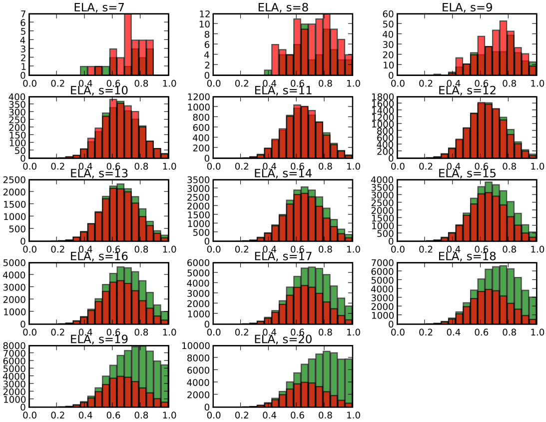



(a) ELA (b) ELM

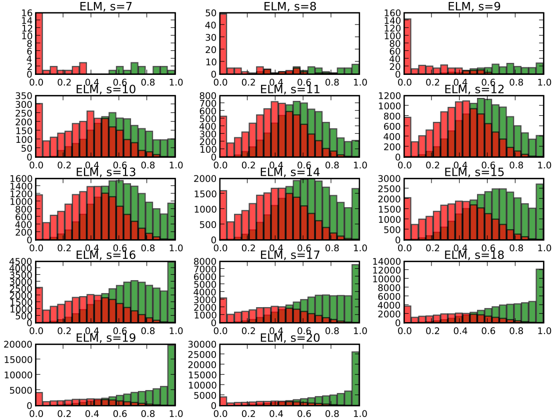
# Restricted Histograms

- sometimes the OK/NOK histograms do not correlate with improvement in the prediction
- for most of the patterns, the classifiers give similar outputs $\Rightarrow$ the OK/NOK separation is irrelevant
- restrict the testing set to "unclear" patterns $U(s)$
- $0 \leq s \leq r$ ... degree of consensus
- for any class, at most $s$ classifiers vote for the class
- OK/NOK histograms restricted to $U(s)$

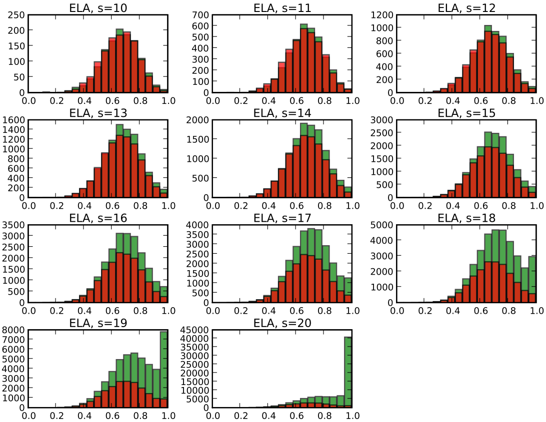
# Restricted Histograms – Waveform



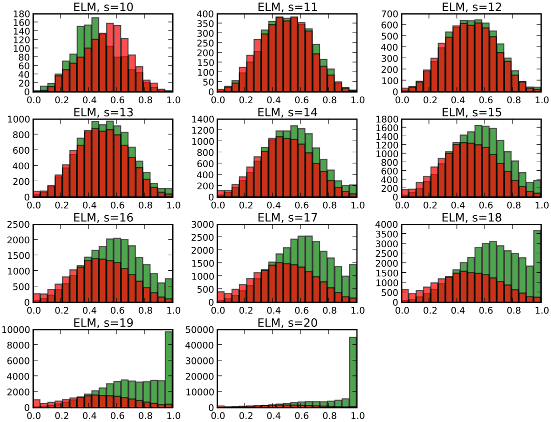
# Restricted Histograms – Waveform



# Restricted Histograms – Phoneme



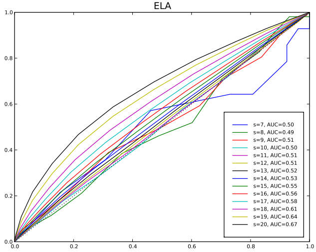
# Restricted Histograms – Phoneme



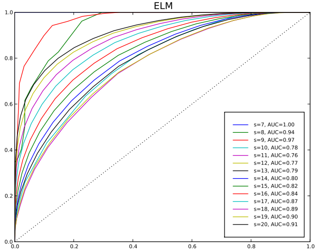
# ROC analysis

- convert the histograms to a ROC curve
- summary – area under ROC (AUC)
- AUC $= 50\%$ – random guessing
- AUC $= 100\%$ – ideal

# OK/NOK ROCs – Waveform

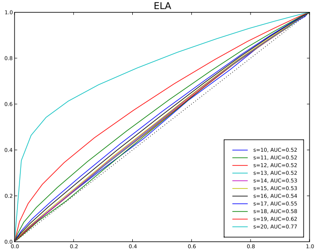



(a) ELA

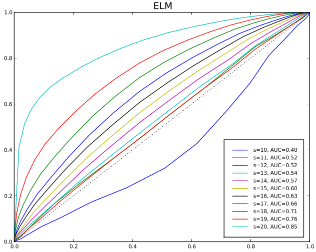



(b) ELM

# OK/NOK ROCs – Phoneme




(a) ELA

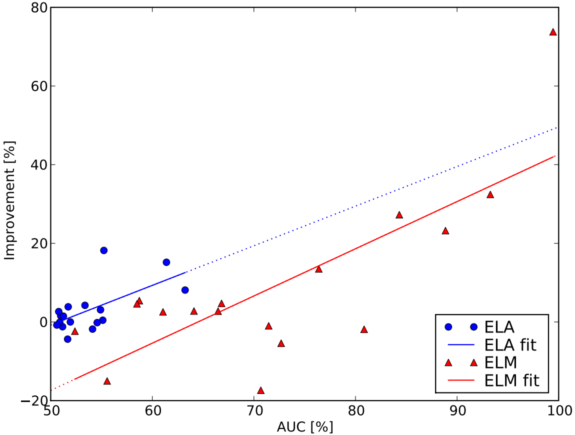



(b) ELM

# Improvement vs. AUC

- AUC for $U(s)$ covering 5% of the data
- improvement of DWM over MV
- better AUC $\overset{?}{\Rightarrow}$ better improvement?
- scatterplot for 17 datasets

# Improvement vs. AUC



# Conclusions & Future Work

- formalism of classifier systems with classification confidence
- 3 types of classifier systems
- experiments with QDCs and RFs: dynamic systems can outperform both static and confidence-free systems
- methods for assessing confidence measures
- future work: different classifier types (SVM)
- future work: more advanced aggregation methods
- better methods for assessing confidence measures

# Thank you for your attention

David Štefka

Dynamic Classifier Systems
for Classifier Aggregation