Dynamic Classifier Systems for Classifier Aggregation

David Štefka

Seminář z umělé inteligence
11. 12. 2008

Martin Holeňa
Outline

1. Classifier Combining
2. Classification Confidence
3. Classifier Systems
4. Experiments
5. Assessing Confidence Measures
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, \ldots)
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, \ldots)
- aggregate the team using aggregator $\mathcal{A}$
- most of the aggregation methods are static
Classification example

- $C_1$ – bananas, $C_2$ – apples
Classification example ctnd.

- $\phi(\vec{x}) = (0.1, 0.8)$
Classification example ctnd.

- $\phi(\bar{x}) = (0.9, 0.1)$
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 $\tilde{x}$ by $\phi$
- ability to answer "I don’t know"
- confidence measure $\kappa_{\phi} : \mathcal{X} \rightarrow [0, 1]$
- 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 $\phi$

$$
\kappa_{\phi}^{(GA)}(\vec{x}) = \frac{\sum_{\vec{y} \in \mathcal{M}} I(\phi_{cr}(\vec{y}) = c(\vec{y}))}{|\mathcal{M}|}
$$
Euclidean Local Accuracy - ELA

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

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

\[
\kappa_{\phi}^{(ELM)}(\vec{x}) = \frac{\sum_{\vec{y} \in N(\vec{x})} I(\phi_{cr}(\vec{x}) = c(\vec{y}))}{|N(\vec{x})|}
\]
Classifier Systems

- $S = (T, K, A)$ – classifier system
- $T = (\phi_1, \ldots, \phi_r)$ – classifiers
- $K = (\kappa_{\phi_1}, \ldots, \kappa_{\phi_r})$ – confidence measures
- $A$ – 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

\[
T(\vec{x}) = \begin{pmatrix}
\phi_1(\vec{x}) \\
\phi_2(\vec{x}) \\
\vdots \\
\phi_r(\vec{x}) 
\end{pmatrix}
= \begin{pmatrix}
\mu_{1,1}(\vec{x}) & \mu_{1,2}(\vec{x}) & \cdots & \mu_{1,N}(\vec{x}) \\
\mu_{2,1}(\vec{x}) & \mu_{2,2}(\vec{x}) & \cdots & \mu_{2,N}(\vec{x}) \\
\vdots \\
\mu_{r,1}(\vec{x}) & \mu_{r,2}(\vec{x}) & \cdots & \mu_{r,N}(\vec{x})
\end{pmatrix},
K(\vec{x}) = \begin{pmatrix}
\kappa_{\phi_1}(\vec{x}) \\
\kappa_{\phi_2}(\vec{x}) \\
\vdots \\
\kappa_{\phi_r}(\vec{x})
\end{pmatrix}
\]
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

![Diagram showing results for QDC]

- **Confidence-free**
- **Static**
- **Dynamic**

<table>
<thead>
<tr>
<th>Combination method</th>
<th>Mean Err ± Std Err</th>
</tr>
</thead>
<tbody>
<tr>
<td>QDC - Breast</td>
<td></td>
</tr>
</tbody>
</table>

- Mean Error (Mean Err ± Std Err) for different combination methods.
## Results – QDC

<table>
<thead>
<tr>
<th>Combination method</th>
<th>Mean Err</th>
<th>Std Err</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence-free</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dynamic</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

David Štefka

Dynamic Classifier Systems for Classifier Aggregation
Results – QDC

![Chart showing Mean Err ± Std Err for QDC Concentric with three methods: Confidence-free, Static, Dynamic. The y-axis represents Mean Err ± Std Err, ranging from 1 to 5. The x-axis represents Combination method, with three categories: Confidence-free, Static, Dynamic. The chart indicates that Dynamic method has the lowest Mean Err ± Std Err compared to Static and Confidence-free methods.](chart.png)
Results – QDC

QDC – Gauss 3D

Mean Err ± Std Err

Confidence-free  Static  Dynamic

Combination method
Results – QDC

![Chart showing QDC Phoneme results with confidence-free, static, and dynamic combination methods. Mean Error ± Standard Error is plotted against combination methods.]
Results – QDC

<table>
<thead>
<tr>
<th>Combination method</th>
<th>Mean Err ± Std Err</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence-free</td>
<td>QDC − Pima</td>
</tr>
<tr>
<td>Static</td>
<td>16</td>
</tr>
<tr>
<td>Dynamic</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>34</td>
</tr>
</tbody>
</table>
Results – QDC

QDC – Satimage

Confidence-free
Static
Dynamic

Mean Err ± Std Err

David Štefka
Dynamic Classifier Systems for Classifier Aggregation
Results – QDC

![Graph showing QDC - Waveform with mean error and standard error for confidence-free, static, and dynamic combination methods.]
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 successful 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 computational 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
sometimes the OK/NOK histograms do not correlate with improvement in the prediction

for most of the patterns, the classifiers give similar outputs ⇒ the OK/NOK separation is irrelevant

restrict the testing set to “unclear” patterns $U(s)$

$0 \leq s \leq r$ \ldots degree of consensus

for any class, at most $s$ classifiers vote for the class

OK/NOK histograms restricted to $U(s)$
Restricted Histograms – Waveform
Classifier Combining
Classification Confidence
Classifier Systems
Experiments
Assessing Confidence Measures

Restricted Histograms – Waveform

ELM, s=7

ELM, s=8

ELM, s=9

ELM, s=10

ELM, s=11

ELM, s=12

ELM, s=13

ELM, s=14

ELM, s=15

ELM, s=16

ELM, s=17

ELM, s=18

ELM, s=19

ELM, s=20
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 $\Rightarrow$ better improvement?
- scatterplot for 17 datasets
Improvevnt vs. AUC

Diagram showing the relationship between Improvement percentage and AUC percentage, with markers for ELA, ELA fit, ELM, and ELM fit, showing a positive correlation. The graph includes a line of best fit for each classifier type.
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