Classification in Presence of Drift and Latency

HaCDAIS Workshop at ICDM2011 / Vancouver, Canada / 2011-12-11

Georg Krempl
Knowledge Management & Discovery
Otto-von-Guericke University Magdeburg
g.krempl@iti.cs.uni-magdeburg.de

Vera Hofer
Statistics & Operations Research
University of Graz
vera.hofer@uni-graz.at
Outline

► Introduction
  ► Motivation
  ► Population Drift
  ► Verification Latency

► Mining Drift: Generative Models
  ► Drifting Decision Boundary
  ► Drifting Subpopulations
  ► Global Prior Drift

► Experiments

► Conclusion
Motivation

**Why study & categorize drift?**

- All adaptive classification models make *assumptions on the type of drift*
- Alignment of adaptive strategy to drift in reality requires *identification* and *categorization* of drift
Population Drift

Changes in distributions over time: Kelly et al., 1999
Here used synonymously to concept drift (Schlimmer and Granger, 1986)

- Static feature space
- Drift can affect:
  - Posterior distribution $P(Y|X)$
  - Feature distribution $P(X)$
  - Class prior distribution $P(Y)$

- Notation:
  - $X$ Explanatory variable(s) (feature(s))
  - $Y$ Binary response (label)
Example

- Classifier for credit scoring:
  Predict default of loan
- Maturity of loan: 3 years
- Most recent available labelled data:
  \textbf{November 2008}

Representative for today's applications?

\textbf{Verification Latency}

Time interval between \textit{classification} and \textit{verification of the prediction} (Marrs et al., 2010)

Also denoted as:
- Time lag (Lucas, 2004)
- Label delay (Kuncheva, 2008)
Concurrence of Drift & Latency

Data Availability Problem:
Whenever predicting outcomes far in the future:
- Available labelled data is outdated
- No actual and labelled data is available
- labelled (old) data
- new (unlabelled) data

Idea of Drift Mining:
- Analyse change in historic, labelled data: Is drift *systematic*?
- Identify invariances in change:
  Do *drift patterns* exist, that relate posterior change to
  - the course of time,
  - changes in the feature distribution?
- Predict current joint and posterior distributions
- Update classifier accordingly
- Drift Models
Drifting Decision Boundary

- Strong relation between $X$ and $Y$, threshold $\tau$ determines class

- Threshold changes over time (cmp. moving hyperplane, Hulten et al.(2001))

- **Drift pattern:**
  Direct relation between posterior drift and course of time

**Approach:**
Learn movement of dec. boundary
Drifting Sub-Populations

- Differently evolving subpopulations
- Clusters evolve gradually

**Drift Pattern:**
Relation between change of $P(Y|X)$ and $P(X)$

**Approach:**
Identify & track sub-populations in unlabelled data over time
Global Prior Drift

- Change of class prior (over the whole feature space)
- Multiplicative model, growth factors $\delta_p, \delta_n$

Drift pattern:
Relation between change of $P(Y|X)$ and $P(X)$

Approach: Estimate $\delta$s from unlabelled data
Can $P(Y)$ be estimated by analysing $P(X)$ in reality?

Results on a real-world credit scoring data set:\footnote{In upcoming publication Mining Drift in Data (contact me for details).}

### True Prior Changes $\delta_p$

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>1.87</td>
<td>2.81</td>
<td>2.54</td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td>—</td>
<td>1.50</td>
<td>1.36</td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>—</td>
<td>—</td>
<td>0.90</td>
<td></td>
</tr>
</tbody>
</table>

### Predicted Prior Changes $\hat{\delta}_p$

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>1.84</td>
<td>2.82</td>
<td>2.52</td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td>—</td>
<td>1.57</td>
<td>1.40</td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>—</td>
<td>—</td>
<td>0.88</td>
<td></td>
</tr>
</tbody>
</table>

Results obtained using a SSE-minimizing estimate of the feature distribution.
Conclusion & Outlook

Drift Mining:

- Aim: Identification of drift patterns (invariances in the change of distributions)
- Use knowledge of drift as substitute for new, labelled data
- Applicable on systematic drift
- Advantageous in presence of verification latency
  Always up-to-date classifier

Current & Future Work:

- Application to more real-world data sets (in different application domains)
- Extension of drift models

Thank you for your attention!

Special thanks to Gary Marrs, Myra Spiliopoulou, Bernhard Nessler and to the referees!

Contact: Georg Krempl
KMD Workinggroup Magdeburg
g.krempl@iti.cs.uni-magdeburg.de
G. Hulten, L. Spencer, and P. Domingos.
Mining time-changing data streams.

M. G. Kelly, D. J. Hand, and N. M. Adams.
The impact of changing populations on classifier performance.

G. Krempl.
The algorithm apt to classify in concurrence of latency and drift.

G. M. Krempl.
Adaptive Prediction Models and their Application to Credit Scoring.

L. I. Kuncheva.
Classifier ensembles for detecting concept change in streaming data: Overview and perspectives.

A. Lucas.
Updating scorecards: Removing the mystique.

G. Marrs, R. Hickey, and M. Black.
The impact of latency on online classification learning with concept drift.

J. C. Schlimmer and R. H. Granger.
Beyond incremental processing: Tracking concept drift.
## Application of Drift Mining

<table>
<thead>
<tr>
<th></th>
<th>No Latency</th>
<th>Latency</th>
<th>Drift Mining</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systematic Drift</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Unsystematic Drift</td>
<td>✓</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Incremental Learning</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>