

Learning in the Presence of Concept Drift and Hidden Contexts

Paper by Gerhard Widmer and Miroslav Kubat, 1996.

Seminar aus Maschinellern Lernen -

FB Informatik, FG Knowledge Engineering, WS 2006/2007

10/01/2007

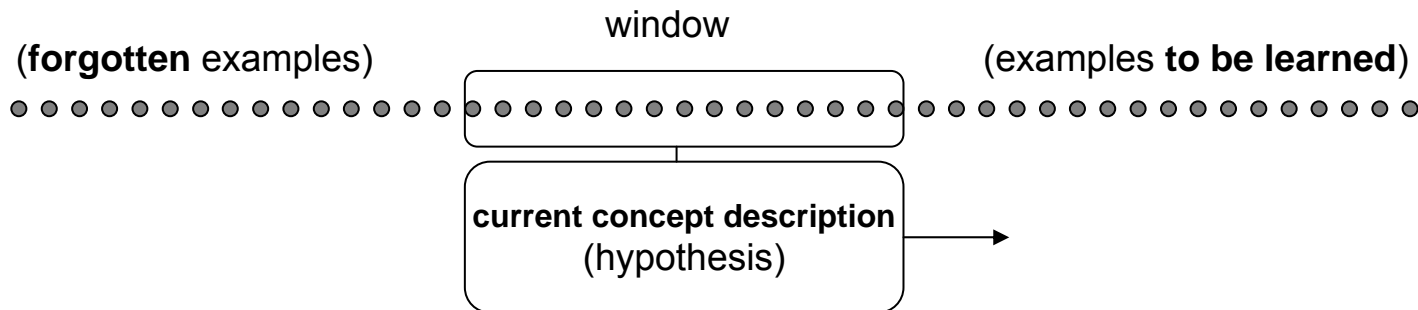
René Moch

Übersicht

- Einleitung
- Algorithmen der FLORA-Familie
 - FLORA Framework, Idee & Funktionsweise
 - FLORA2, Konzeptverschiebung (*concept drift*)
 - FLORA3, Versteckte Kontexte (*hidden contexts*)
 - FLORA4, Rauschen in den Daten (*drift vs. noise*)
- Zusammenfassung

Einleitung

- Inkrementelles (bzw. online) Lernen von Konzepten
- „*Learning from Examples*“
- Relevante Daten als *object stream*



- Annahme: nur aktuelle Instanzen (Objekte) besitzen Relevanz, ältere werden „vergessen“

Einleitung

- Instanzen (Objekte) besitzen (*o.B.d.A.*) binäre Klassifikation c mit $c \in \{\text{positive, negative}\}$
- Prozess: *sequence of trials* -> **Beschreibung des Zielkonzeptes**
 1. *Learner* erhält Instanz von *Teacher*
 2. *Learner* klassifiziert Instanz als positiv oder negativ
 3. *Learner* erhält Feedback von *Teacher*, ob Klassifikation der Instanz korrekt war
 4. *Learner* nutzt Feedback zum Update seiner Konzeptbeschreibung

Einleitung

- Ziel: Stabile Konzeptbeschreibung (*Hypothese*) zur Vorhersage von Objektklassifikationen
- Minimierung der Anzahl von Fehlklassifikationen (in der Vorhersage)
- Probleme:
 - *Concept Drift*
 - *Hidden Contexts*

General FLORA Framework: Logik

- FLORA Algorithmen basieren auf einfacher Attribut-Wert-Logik ohne Negation
- Beschreibende Elemente: Deskriptoren (*description items*) (=Konjunktionen von AW-Paaren)
- Beispiele:
 - (color = white) AND (temperature = low) ist TRUE für „SNOW“
 - (shape = cube) ist FALSE für „GLOBE“

General FLORA Framework: Concept Description

- Hypothese repräsentiert durch 3 Deskriptormengen (*description sets*):
 - *ADES* = $\{ADes_1/AP_1, ADes_2/AP_2, \dots\}$
 - *Accepted DEScriptors*, nur positive Instanzen
 - *NDES* = $\{NDes_1/NN_1, NDes_2/NN_2, \dots\}$
 - *Negative DEScriptors*, nur negative Instanzen
 - *PDES* = $\{PDes_1/PP_1/PN_1, \dots\}$
 - *Potential DEScriptors*, zu allgemein, daher sowohl positive als auch negative Instanzen („*Deskriptor-Reservoir*“)

General FLORA Framework: Concept Description Process

- Verschieben des Fensters bewirkt Addition und Löschen jeweils eines Objektes aus der Hypothese
- *ADES* und *NDES* durch inkrementelle Generalisierung erzeugt (Gen. = Entfernen einzelner AW-Paare aus Konjunktion eines Deskriptors)
- *PDES* enthält verschobene Deskriptoren, die einst in *ADES* oder *NDES* waren
- Update relevanter Zähler
 - Addition pos. (neg.) Objekte
 - Löschen pos. (neg.) Objekte

General FLORA Framework: Basis Algorithmus: *Learning*

```
Function learn_from(I) für eine positive Instanz I
algorithm:
MATCH := false;
for i := 1 to |ADES|
  if match(I, ADesi) then
    begin   APi := APi + 1;
           MATCH := true;
    end;
if not MATCH then G := generalize(I, ADES, PDES, NDES);
if not MATCH and not G then include(I/1, ADES);
for i := 1 to |PDES|
  if match(I, PDesi) then PPi := PPi + 1;
for i := 1 to |NDES|
  if match(I, NDesi) then
    begin   delete(NDesi, NDES);
           include(NDesi/1/NNi, PDES);
    end;
end;
```

General FLORA Framework: Basis Algorithmus: *Forgetting*

Function `forget(I)` für eine positive Instanz I

algorithm:

```
for  $i := 1$  to  $|ADES|$ 
  begin if  $match(I, ADes_i)$  then  $AP_i := AP_i - 1$ ;
        if  $AP_i = 0$  then  $delete(ADes_i, ADES)$ ;
  end;
for  $i := 1$  to  $|PDES|$ 
  begin if  $match(I, PDes_i)$  then  $PP_i := PP_i - 1$ ;
        if  $PP_i = 0$  then
          begin  $delete(PDes_i, PDES)$ ;
                 $include(PDes_i/PN_i, NDES)$ ;
          end;
  end;
end;
```

Flexible Windowing: FLORA2

- Problem: Concept Drift
- Analyse: Fenstergröße
 - zu klein: keine ausreichend hohe Anzahl von Instanzen für stabile Konzeptbeschreibung, Umfang der Hypothese zu gering
 - zu groß: Reaktionszeit auf Konzeptverschiebung zu hoch, Anzahl der Fehlklassifikationen steigt dramatisch
- FLORA2: dynamische Anpassung der Fenstergröße
- Ausmaß der Konzeptverschiebung bestimmt die optimale Einstellung der Fenstergröße!



Flexible Windowing: FLORA2

Heuristische Indikatoren

1. **System Akkuratheit** *Acc* (= Vorhersage Performanz gemessen an Erfolgsrate in fester Anzahl vergangener Klassifik.)
2. **Syntaktische Komplexität** (der entwickelten Hypothese)
 - Annahme: Concept Drift, falls:
 - drastischer Abfall von *Acc* (siehe 1.)
 - explosionsartiger Anstieg der Anzahl von Deskriptoren in *ADES* (siehe 2.)

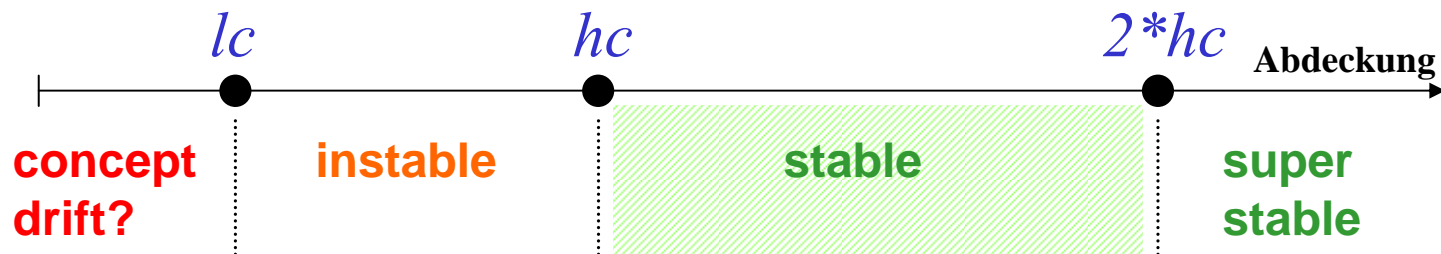


Flexible Windowing: FLORA2

Anpassung der Heuristik

- 3 Parameter zur Anpassung an die Anwendung
 - lc (= Grenzwert für niedrige Abdeckung von *ADES*)
 - hc (= Grenzwert für hohe Abdeckung von *ADES*)
 - p (= Grenzwert für akzeptable Akkuratheit der Vorhersage)

$$\text{Abdeckung} = \frac{\# \text{ durch Deskriptoren abgedeckte Instanzen}}{\# \text{ hierfür benötigte Literale}}$$





Flexible Windowing: FLORA2

Window Adjustment Heuristic

denotations:

N Anzahl positiver Instanzen abgedeckt durch ADES
 S Größe von ADES in Anzahl verwendeter Literale
 Acc gegenwärtige Akkuratheit der Vorhersage (gemessen an Erfolgsrate der letzten Klassifikationen)

|W| Fenstergröße

Parameter lc, hc und p sind User definiert.

algorithm:

```

if (N/S < lc) OR ((Acc < p) AND decreasing(Acc)) /*drift suspected*/
  then L := 0.2*|W|                               /*reduce window by 20%*/
else if (N/S > 2*hc) AND (Acc > p)               /*extremely stable*/
  then L := 2                                     /*reduce window by 1*/
else if (N/S > hc) AND (Acc > p)                 /*stable enough*/
  then L := 1                                    /*keep window fixed*/
else L := 0                                       /*grow window by 1*/
  
```



Flexible Windowing: FLORA2

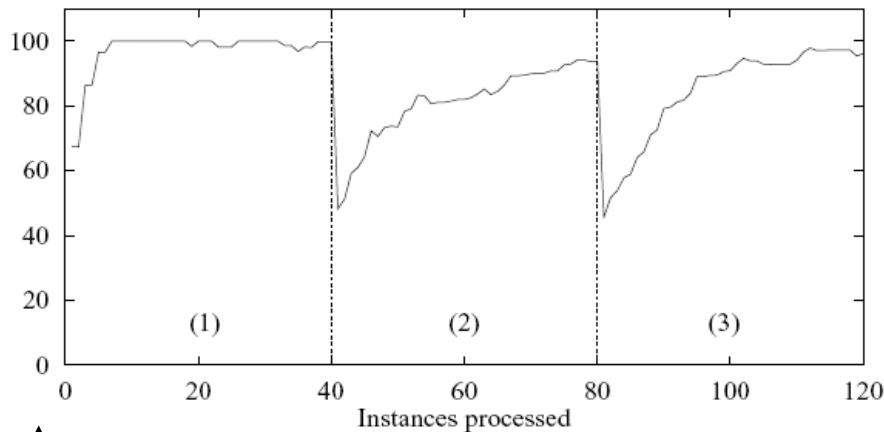
The *STAGGER* concepts

- Experimente mit „*STAGGER* concepts“ rechtfertigen folgende Parameter Konfiguration: $lc = 1.2$; $hc = 4.0$; $p = 70\%$
- Instanzraum
 - size ε {small, medium, large}
 - color ε {red, green, blue}
 - shape ε {square, circular, triangular}
- Ziel Konzepte
 - (1) size = small \wedge color = red
 - (2) color = green \vee shape = circular
 - (3) size = (medium \vee large)

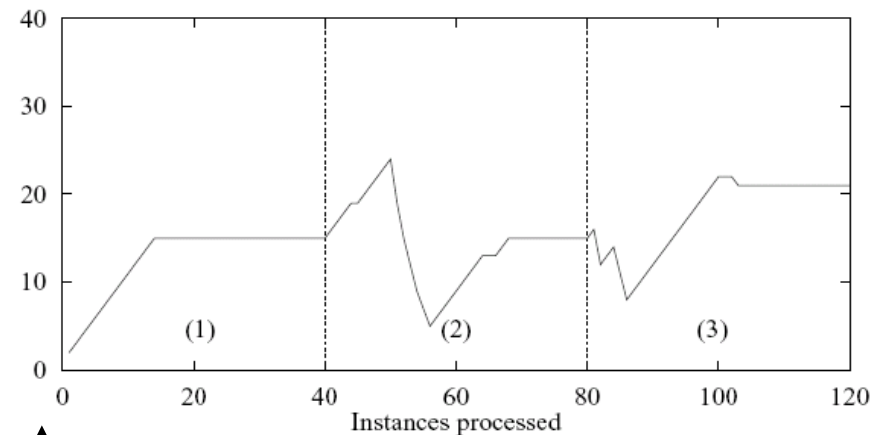
Flexible Windowing: FLORA2

The *STAGGER* concepts

- 120 zufällige Trainingsinstanzen, Concept Drift nach jeweils 40 Instanzen
- 100 zufällige Testinstanzen, Ergebnis über 10 Durchläufe gemittelt



Akkuratheit der Vorhersage in %



Fenstergröße in #Instanzen

Window Adjusting Heuristic (WAH)

Dealing with Recurring Contexts: FLORA3

- Problem: Hidden Contexts
 - (Biologische und ökonomische) Systeme durchlaufen Zyklen
 - Systeme entwickeln sich durch wiederkehrende Muster
 - z.B.: 4 Jahreszeiten
- Konzepte immer wieder **neu** lernen?
- Lösung: Stabile Konzeptbeschreibungen speichern und später bei Bedarf auf aktuelle Relevanz prüfen
- FLORA3: storage & recall Mechanismus

Dealing with Recurring Contexts: FLORA3

- **Storage: was?**
 - Konzeptbeschreibung als Tripel $\{ADES, PDES, NDES\}$ repräsentiert und auch als solches gespeichert
 - Zähler irrelevant
- **Storage: wann?**
 - Bed.: Konzept stabil gemäß WAH
 - Bed.: kein identisches *ADES* bereits im Speicher vorhanden
- **Recall: was?**
 - Choose context ...
- **Recall: wann?**
 - Bed.: WAH vermutet Konzeptverschiebung (Reduktion der Fenstergröße um 20%)



Dealing with Recurring Contexts: FLORA3

Choose Context

1. Find best candidate

- Kandidat, falls konsistent mit aktueller Instanz
- Entscheidend: Quotient aus positiven und negativen Instanzen, die gematcht werden

2. Update the best candidate

- Alle Instanzen im Fenster regeneralisieren
 - Alle Zähler in den description sets auf *Null* setzen
 - Alle Instanzen im Fenster „neu lernen“ (FLORA *Learning*)

3. Compare the updated best candidate

- $C_{\text{updated_best}}$ besser als C_{current} ?
- Relevant: relative Komplexität der Beschreibung (bestes Konzept besitzt **prägnanteste Menge ADES**)



Dealing with Recurring Contexts: FLORA3

Choosing Context Process

denotations:

STABLE Boolesche Variable, TRUE falls aktuelle Hypothese stabil gemäß WAH

DRIFT_SUSPECTED Boolesche Variable, TRUE falls WAH einen concept drift vermutet und das Fenster verkleinert hat

functions:

store_current Speichere aktuelle description sets (ident. *ADES*?)

find_best_candidate Finde besten gespeicherten „matching“ context

regeneralize_old_description Regeneralsiere Instanzen im aktuellen Fenster

replace_if_applicable Stelle alte Hypothese wieder her, falls besser als aktuelle

Function `choose_context`

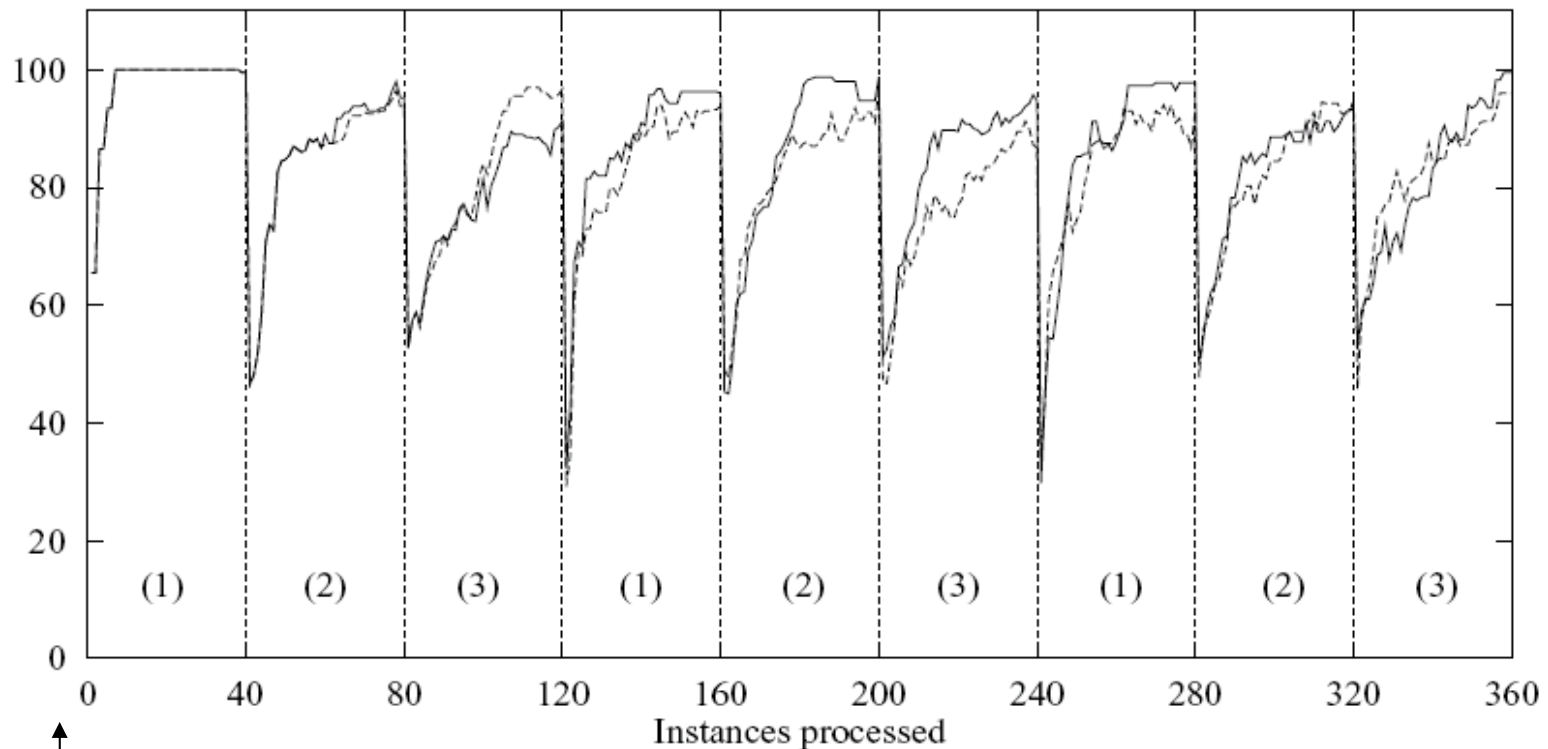
algorithm:

```
if STABLE then store_current;
else if DRIFT_SUSPECTED then
  begin   Best := find_best_candidate;
         G := regeneralize_old_description(Best);
         replace_if_applicable(G);
  end;
```



Dealing with Recurring Contexts: FLORA3

FLORA3 vs. FLORA2 on *STAGGER* concepts



Akkuratheit der Vorhersage in % / FLORA3 (*solid lines*), FLORA2 (*dashed lines*), cyclic order of *STAGGER* concepts

Drift vs. Noise: FLORA4

- Problem: Destabilisierung bei sehr langsamer Konzeptverschiebung und Rauschen (*noise*)
- *Concept drift* oder „nur Irregularität“?
- Vorteil bisher: Schnelle Reaktion auf *Drift*
- Nachteil jetzt: Überreaktion bei *Noise*
- Ursache hierfür: strikte Konsistenzbedingungen für *ADES*
- FLORA4: Kompromisslösung mit Konfidenzintervallen (statistisches Rating der Deskriptoren)



Drift vs. Noise: FLORA4

Konfidenzintervalle für Deskriptoren

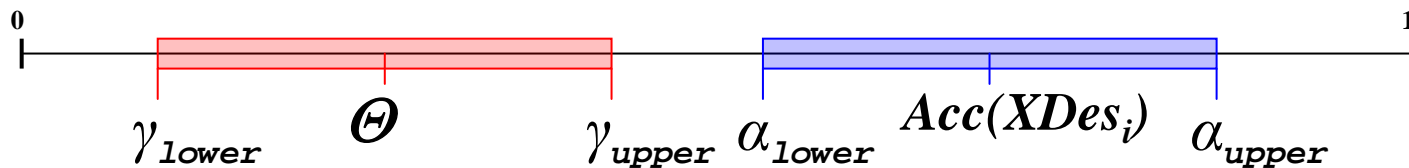
- **Deskriptor Akkuratheit** $Acc(XDes_i) \in [0, 1]$ mit den Schranken α_{lower} und α_{upper} zu einem bestimmten Konfidenzniveau μ
 - berechnet über Instanzen des aktuellen Fensters
- **Relative Häufigkeit der von $XDes_i$ vorhergesagten Klasse** $\Theta \in [0, 1]$ mit den Schranken γ_{lower} und γ_{upper} zum selben Konfidenzniveau μ
 - berechnet über bislang beobachtete Instanzen
- Relative Position der Konfidenzintervalle für Parameter $Acc(XDes_i)$ und Θ zueinander?



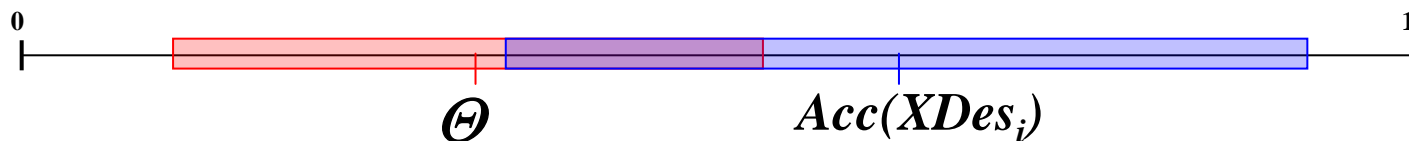
Drift vs. Noise: FLORA4

Beispiele: Konfidenzintervalle

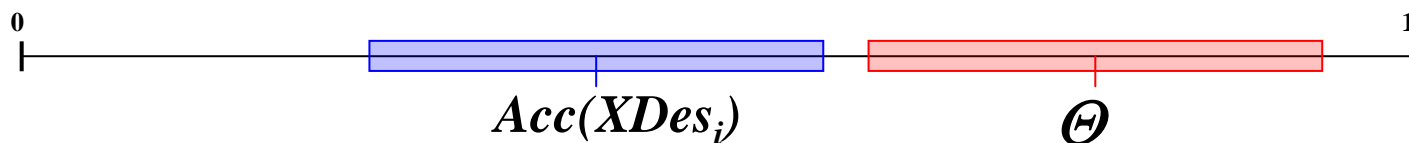
Szenario 1 (Bsp.)



Szenario 2 (Bsp.)



Szenario 3 (Bsp.)



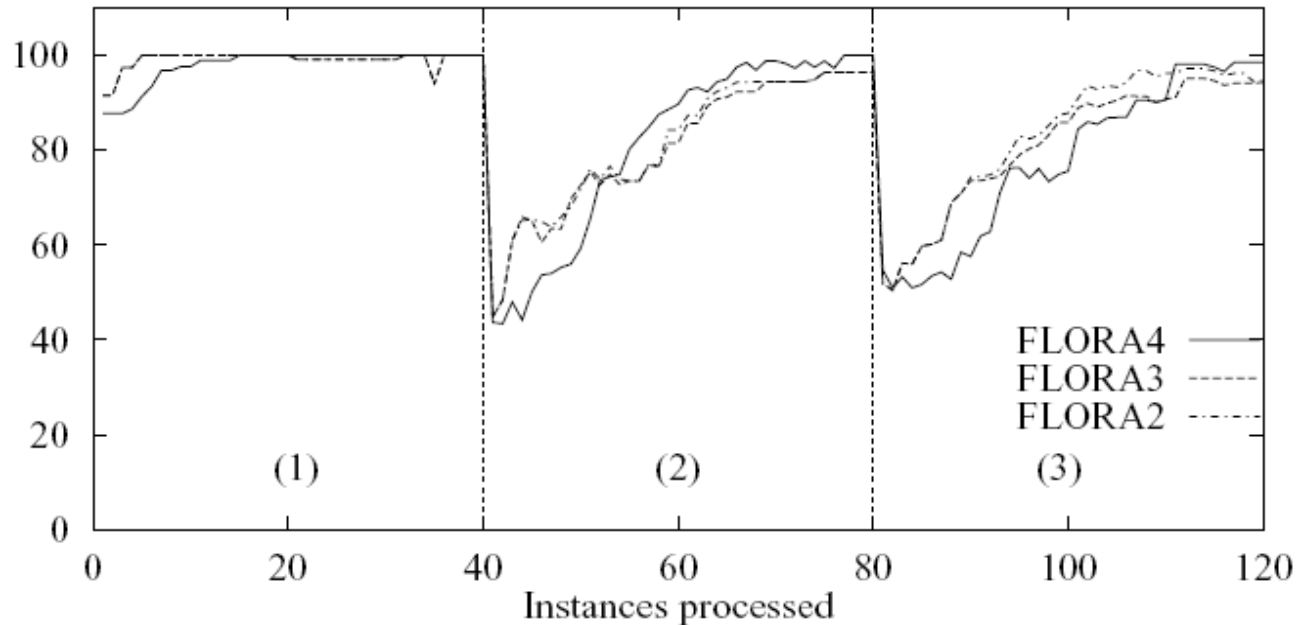
Drift vs. Noise: FLORA4

Maintaining the Description Sets

- Szenario 1: Deskriptor X verbleibt in *ADES*
 - $(\alpha_{\text{lower}} > \gamma_{\text{upper}})$
 - dann auch X von *PDES* nach *ADES*
 - X wird **temporär akzeptierter** Prediktor
- Szenario 2: X von *ADES* nach *PDES* (überlapp. Interv.)
 - $(\alpha_{\text{upper}} \geq \gamma_{\text{lower}})$ AND $(\alpha_{\text{lower}} \leq \gamma_{\text{upper}})$
 - X ist **unbedeutender** Prediktor (sehr große Konfidenzintervalle)
- Szenario 3: X wird gelöscht
 - $(\alpha_{\text{upper}} < \gamma_{\text{lower}})$
- Szenario 1: Deskriptor Y verbleibt in *NDES*
 - $(\alpha_{\text{lower}} > \gamma_{\text{upper}})$
 - Achtung: Kein Verschieben von Y aus *NDES* nach *PDES*, inakzeptable Y in *NDES* werden gelöscht

Drift vs. Noise: FLORA4

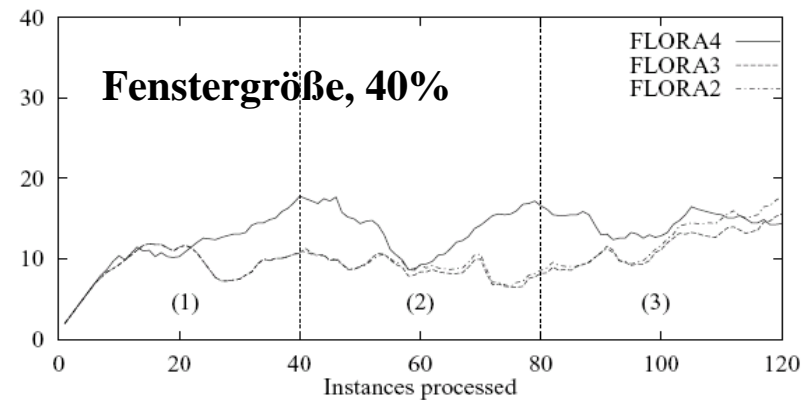
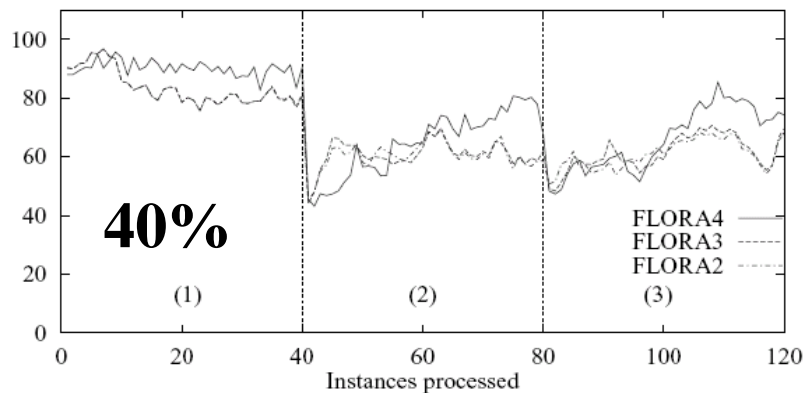
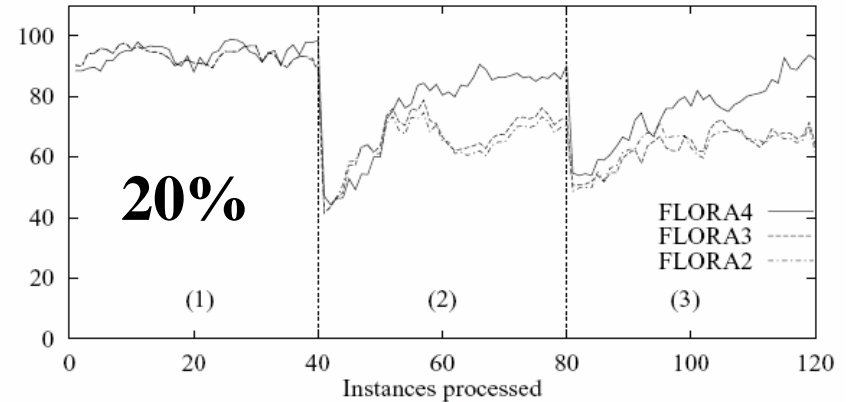
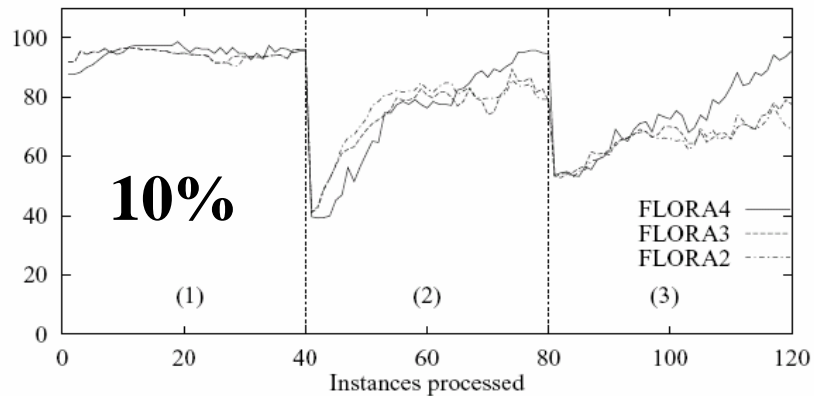
FLORA4 vs. FLORA3 and FLORA2 on *STAGGER* concepts



↑ Akkuratheit der Vorhersage in % / *noise-free drift tracking task*

Drift vs. Noise: FLORA4

FLORA4, FLORA3, FLORA2 at 10%, 20%, 40% Noise



Zusammenfassung

- FLORA: online Learning bei
 - Konzeptverschiebung
 - Kontext-abhängigen Konzepten
 - Konzeptverschiebung & Noise
- Systemschwäche: WAH sensibel gegenüber syntaktischer Komplexität
 - Fenstergröße?
 - Konzept stabil?
 - Konzeptverschiebung vermutet?
- Ausblick: FLORA4 mit integrierter *Expectation* (Kontexte in Zyklen)

Quellen

[WK1996] Widmer, Gerhard; Kubat, Miroslav: „**Learning in the Presence of Concept Drift and Hidden Contexts**“, in: *Machine Learning*, 23, pp. 69-101, 1996.

[HSH1998] Harries, Michael; Sammut, Claude; Horn, Kim: „**Extracting Hidden Context**“, in: *Machine Learning*, 32, pp. 101-126, 1998.

[KW1995] Kubat, Miroslav ; Widmer, Gerhard : „**Adapting to Drift in Continuous Domains**“, *Lecture Notes in Computer Science*, 912, p. 307 et seqq., 1995.

[WG2002] Wang, Jun; Gasser, Les: „**Mutual Online Concept Learning for Multiple Agents**“, *International Conference on Autonomous Agents, Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1*, pp. 362-369, 2002.

[Wid1993] Widmer, Gerhard: „**Effective Learning in Dynamic Environments by Explicit Context Tracking**“, *European Conference on Machine Learning, Proceedings*, 667, pp. 227-243, 1993.

[Wid1994] Widmer, Gerhard: „**Combining Robustness and Flexibility in Learning Drifting Concepts**“, *European Conference on Artificial Intelligence*, pp. 468-472, 1994.

[WK1992] Widmer, Gerhard; Kubat, Miroslav: „**Learning Flexible Concepts from Streams of Examples: FLORA2**“, *European Conference on Artificial Intelligence*, pp. 463-467, 1992.

Fragen & Diskussion.

Learning in the Presence of Concept Drift
and Hidden Contexts