Adaptive Learning from Evolving Data Streams

Albert Bifet and Ricard Gavaldà

Laboratory for Relational Algorithmics, Complexity and Learning, LARCA
Departament de Llenguatges i Sistemes Informàtics
Universitat Politècnica de Catalunya

Intelligent Data Analysis (IDA'09), Lyon, august 2009

Data Streams

Data Streams

- Date items arrive in sequence
- Very long or infinite: Memory sublinear in # elements
- High speed of arrival: low processing time per item
- Anytime answers; OK if approximate

Time Change

- Distribution or concept change
- Abrupt change or gradual change (drift)

Mining Data Streams with Concept Drift

Extract information from

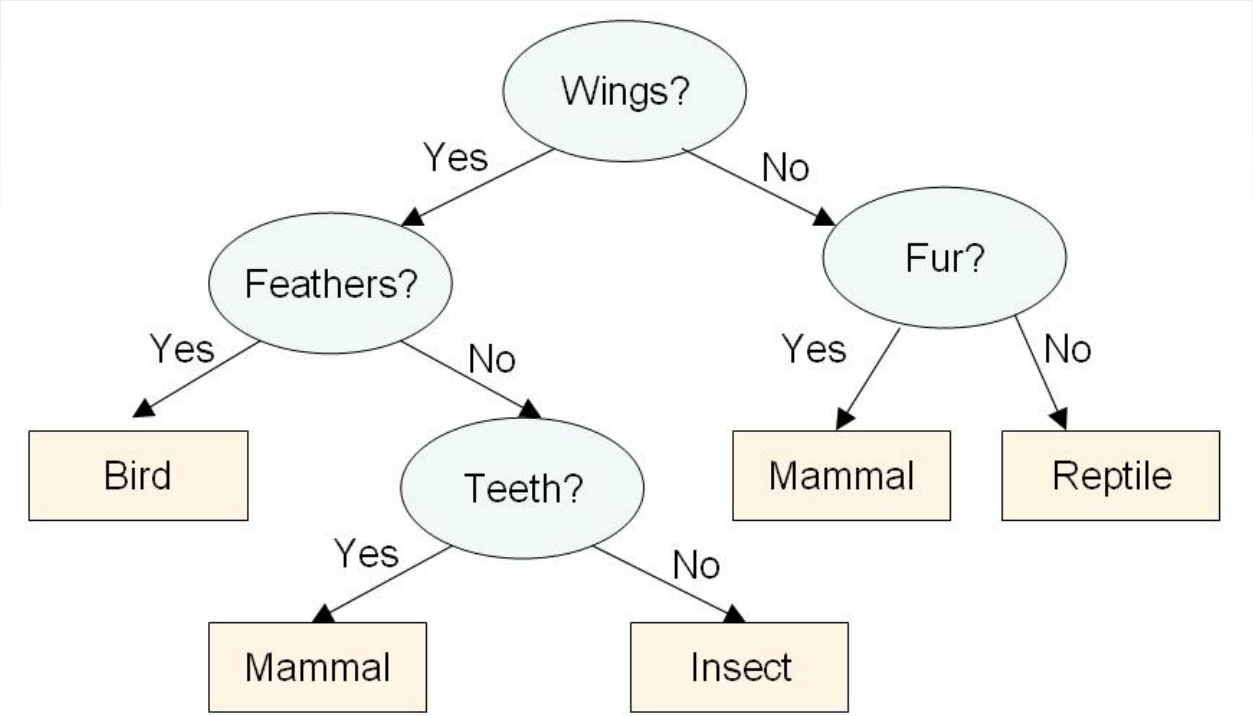
- potentially infinite sequence of data
- possibly varying over time
- using few resources

Adaptively:

- no prior knowledge of type or rate of change
- ignorance does not penalize performance (much)

Decision Trees

- Widely used classifiers; many induction algorithms
- (Here) Discrete attributes and class



VFDT

Very Fast Decision Trees [Domingos,Hulten KDD'00]
Stream-like, but does not handle change

1. For every example x in stream do
2. Sink x to its leaf
3. Update counters for (attribute,value) at leaf
4. **if** counter statistics tell "good enough" attribute a
5. split leaf into node by a

VFDT + change over time

- From time to time, see if change occurs at a node
- If so, create alternative tree
- From time to time, check if alternative better than current

CVFDT [Hulten,Spencer,Domingos KKD'01]

- Concept-adapting Very Fast Decision Trees
- Parameters T_0, T_1, T_2 quantify "from time to time"
- Window W of examples to label current leaves
- Best values depend on rate of change!

Adaptive Hoeffding Tree

- Monitor error rate at each node + change detector
- Keep updated statistics since last change

- “From time to time”: when change detector says so!
- Rebuild tree: from current statistics - no window

ADWIN, *AD*aptive *WIN*dowing

- Change management primitive [Bifet-G SDM'06]
- Window of variable, optimal length W
- Memory, time $O(\log W)$
- Change detection, CUSUM-like
- Estimate of average, EWMA-like
- *No apriori assumptions on rate or nature of change*
- Rigorous guarantees (theorem)

Adaptive Hoeffding Treevs. CVFDT

- Slightly slower - continuous monitoring

+ No apriori knowledge of change rate – parameterless
+ As accurate as CVFDT with *best* setting of CVFDT
+ Much more accurate than CVDT with *wrong* parameters
+ Less memory (no window) for realistic cases
+ Theoretical guarantees
+ Cleaner algorithm

In perspective

Adaptive and parameter-free methods by

- replace frequency statistics counters by `ADWIN`
- encapsulate change detection and estimation in single module
- Parameterless, adaptive, accurate, clean algorithms

Instances

- Naive Bayes, *k*-means clustering [Bifet-G, DS'06, SDM'07]
- Frequent tree mining [B-G, KDD'08, ECMLPKDD'09]
- Ensemble methods [B-H-P-K-G, KDD'09]
- Decision trees [Bifet-G, IDA'09]