Tackling Concept Drift by Temp Inductive Transfer

George Forman
Hewlett-Packard Labs
Presented by Steve Spagnola

Concept Drift

Class Distribution
Subclass distributions change
Hepatitis A: Outbreaks occur with time invariant symptoms

Subclass Distribution - Examined in Paper

Fickle Concept Drift

Cases may change ground truths over time

Concept Drift

Daily Classification Task
Time discretized into days
4 Binary Classification Tasks

Reuters RCV1
Only use top 100 words for each day

Bi-Normal Separation
Like Information Gain

Inductive Transfer

Machine Learning

$h(X_0, X_1, \ldots, X_n) = Y$

Statistical relationship changes over time

$h(X_0, X_1, \ldots, X_{t+1} - Y_{t+1})$
$h(X_0, X_1, \ldots, X_{t+m} - Y_{t+m})$
Temporal Inductive Transfer (TIT) Model

- Use learned models from the past
- Only use yesterday's model (which intrinsically uses
  recurrence relation broken in this paper)
- What would past models predict for today?
- Use learned models from the past

Hence, only one $p$ value

$\{0,1\}$

Figure 2. Temporal Inductive Transfer (TIT) Model

- $P=2$ additional features for each case
- $T=100$ daily training cases
- $T_0$ daily classifier
- $T=100$ daily training cases
- $T_0$ day classifier
- $T=100$ daily training cases
- $T_0$ day classifier

Case study: GCT (governmental social issues; 729 top predictive words)
Results

Conclusions

Effectively tackles concept drift

Hindsight data required for Temporal Induction

Recurrence Chain Broken in tests

Transfer to work

Future Work: increased accuracy if chain preserved