Models for Drift-Adaptive Scoring

Georg Krempl[1], Vera Hofer[1], Mads Krogh Nielsen[2], Troels Verge[2]

[1] University of Graz, Georg.Krempl@uni-graz.at
[2] Danish Ministry of Taxation, Mads.Krogh.Nielsen@skat.dk

2011-08-24

Outline

- Task: **Classification model** (default/non-default) for SKAT Denmark

Outline

- Task: **Classification model** (default/non-default) for SKAT Denmark
- Methodological Problem:
  Classification in concurrence of drift and latency

Outline

- Task: **Classification model** (default/non-default) for SKAT Denmark
- Methodological Problem:
  Classification in concurrence of drift and latency
- Approach:
  Drift mining using explicit drift models
  - Assessing and addressing drift: General approach
  - Explicit drift models (selection)
  - Results on real-world data

Outline

- Task: **Classification model** (default/non-default) for SKAT Denmark
- Methodological Problem:
  Classification in concurrence of drift and latency
- Approach:
  Drift mining using explicit drift models
  - Assessing and addressing drift: General approach
  - Explicit drift models (selection)
  - Results on real-world data
- Related work:

  Machine Learning
  - Semi-supervised learning (but: no drift)
  - Change mining: Böttcher et al., 2008 (but: no latency)

  Credit Scoring
  - Effects of recessions on credit scoring systems: Hoyland, 2004 (overview, ex-post)
  - Effects of 1989-1990 economic turning point on a UK scoring model:
    Crook et al., 2004 (classifier, ex-post)

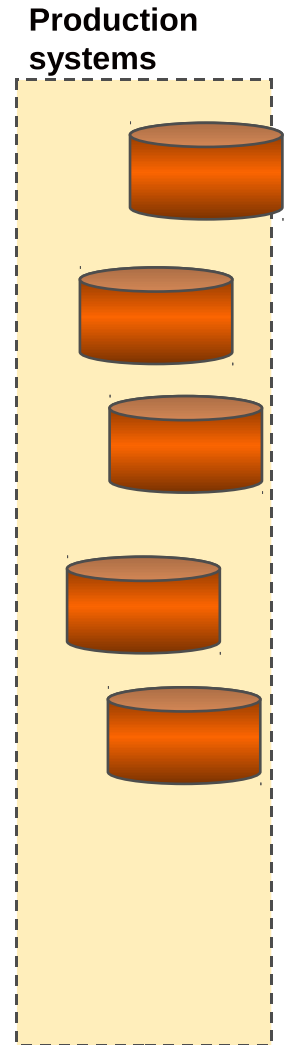
Ministry of Taxation, Denmark – a typical collector's office

Ministry of Taxation, Denmark – a typical collector's office

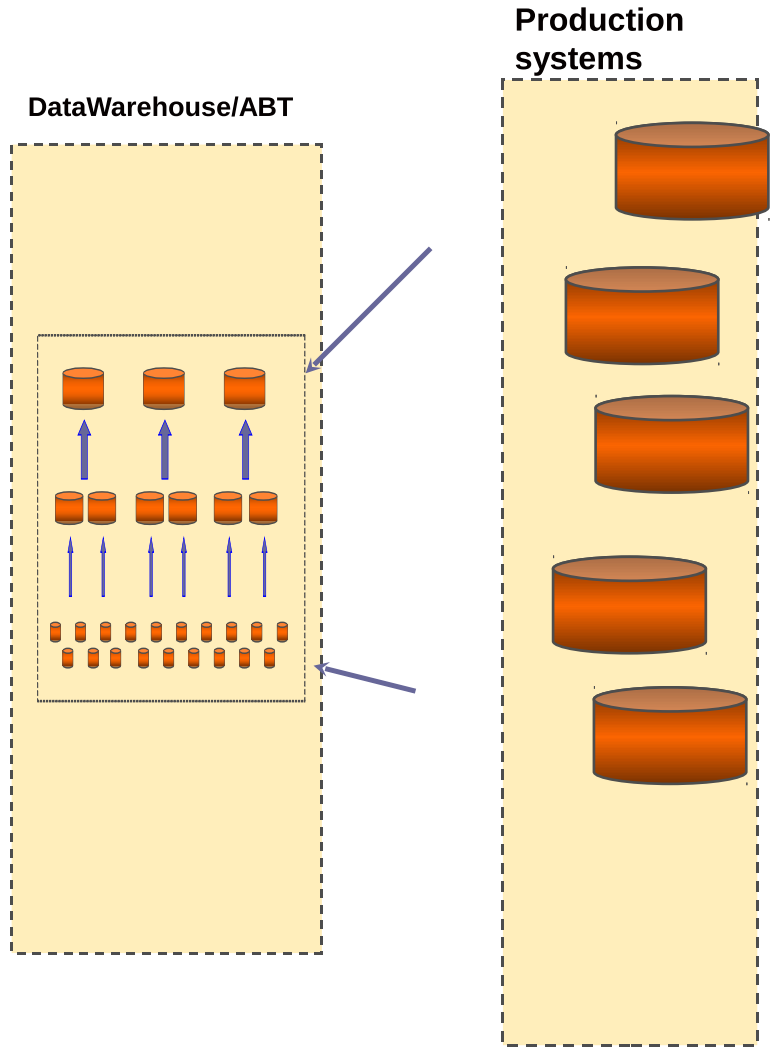
Ministry of Taxation, Denmark – a typical collector's office

- Building a fully automatized collection engine to address all debt to the public sector.

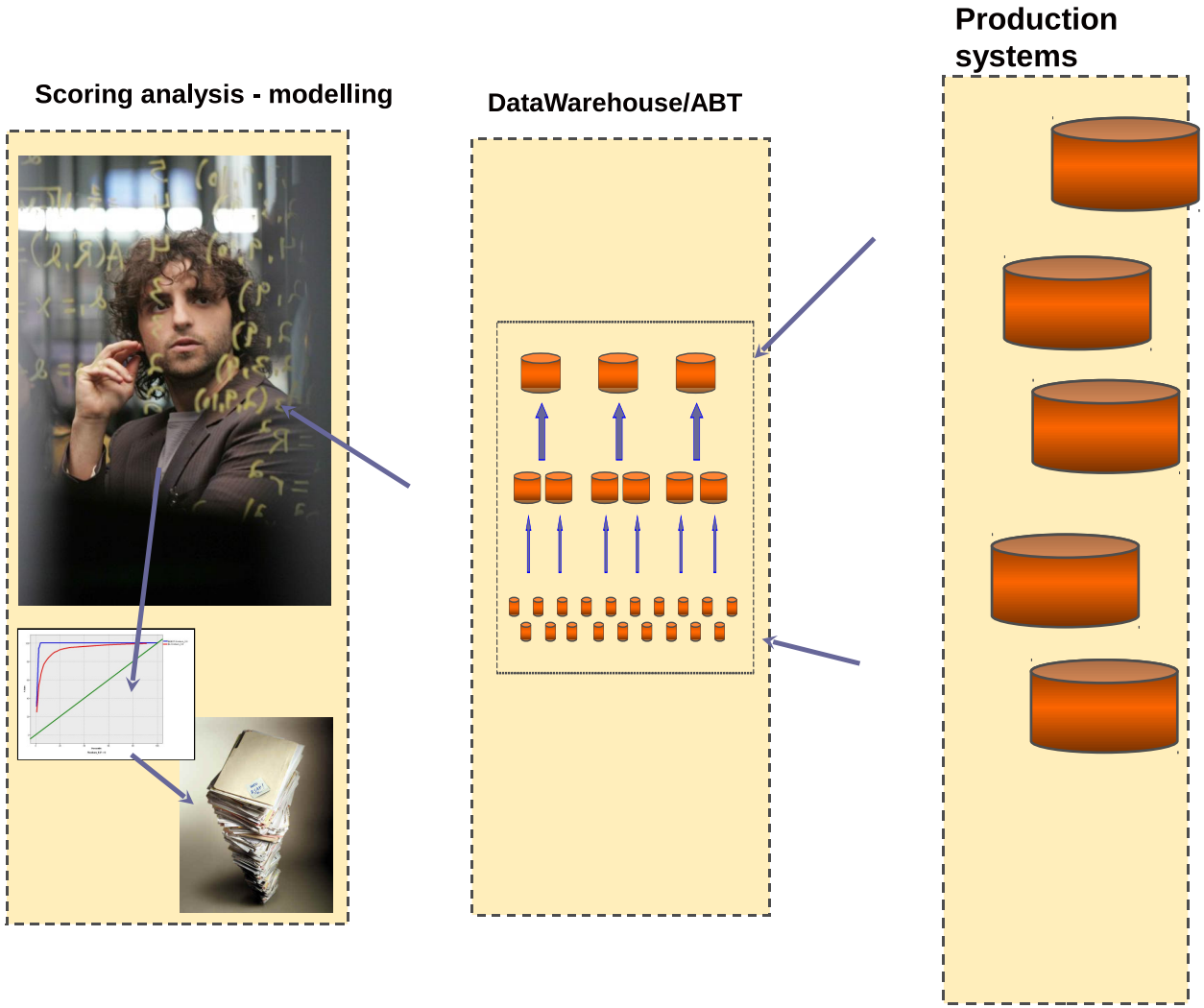
Ministry of Taxation, Denmark – a typical collector's office



Ministry of Taxation, Denmark – a typical collector's office

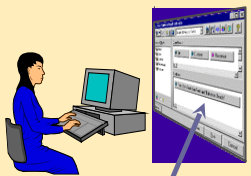


Ministry of Taxation, Denmark – a typical collector's office

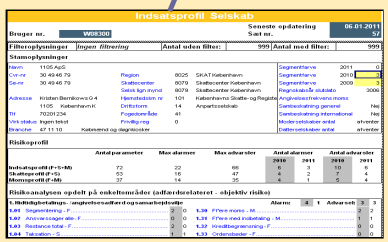


Ministry of Taxation, Denmark – a typical collector's office

Manual procedures being supported

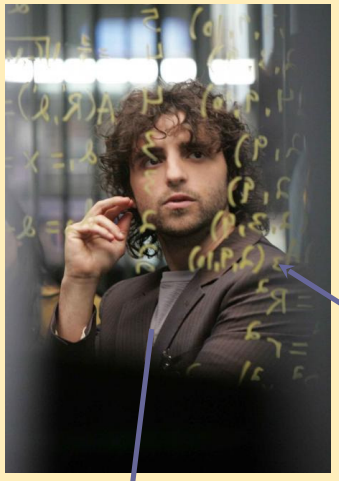


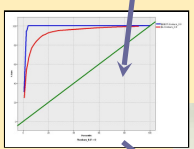
Manual Collection



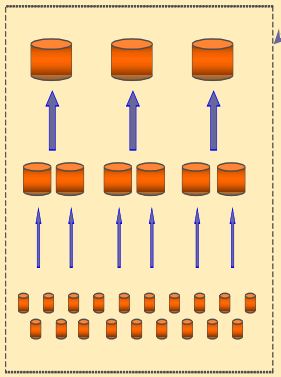
Automatized assessment and collection

Scoring analysis - modelling

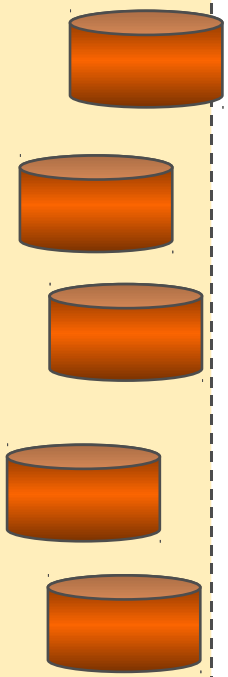




DataWarehouse/ABT



Production systems



Ministry of Taxation, Denmark – a typical collector's office

- Building a fully automatized collection engine to address all debt to the public sector.

Ministry of Taxation, Denmark – a typical collector's office

- Building a fully automatized collection engine to address all debt to the public sector.
- "From cradle to grave"
  - Visitation of the arrears
  - Creation of tracks
  - Planning of schedules

Ministry of Taxation, Denmark – a typical collector's office

- Building a fully automatized collection engine to address all debt to the public sector.
- "From cradle to grave"
  - Visitation of the arrears
  - Creation of tracks
  - Planning of schedules
- 265 types of arrears are treated in the system.
- 2700 variables on persons and companies
- Full population of customers from 2006 to 2009
- Visitation through a selection score based on a LR model on all 265 types of arrears.

- . . . so we already have the data set

What's in it for us?

- 2700 variables on all persons and companies owing money to the public sector 2006-2009
- Complete through the door sample
- No selection sample bias
- No Reject Inference issue
- No bias through marketing etc.

What's in it for us?

- 2700 variables on all persons and companies owing money to the public sector 2006-2009
- Complete through the door sample
- No selection sample bias
- No Reject Inference issue
- No bias through marketing etc.

- Building these models are rather expensive – an investment
  Drift adaptive scoring brings down expenses

Drift and Latency

Population Drift

- Changes in distributions over time: Kelly et al., 1999
- Notation: X explanatory variable, Y binary response
- Effects $P(Y|X)$, $P(X, Y)$, $P(Y)$, $P(X)$

Drift and Latency

Population Drift

- Changes in distributions over time: Kelly et al., 1999
- Notation: X explanatory variable, Y binary response
- Effects $P(Y|X)$, $P(X, Y)$, $P(Y)$, $P(X)$

Verification Latency:

Time interval between *classification*
and *verification of the prediction* (Marrs et al., 2010)

Also denoted as: Time lag (Lucas, 2004), Label delay (Kuncheva, 2008)

Drift and Latency

Population Drift

- Changes in distributions over time: Kelly et al., 1999
- Notation: X explanatory variable, Y binary response
- Effects $P(Y|X)$, $P(X, Y)$, $P(Y)$, $P(X)$

Verification Latency:

Time interval between *classification*
and *verification of the prediction* (Marrs et al., 2010)

Also denoted as: Time lag (Lucas, 2004), Label delay (Kuncheva, 2008)

Concurrence of Drift and Latency:

- No actual and labelled data
- Available old and labelled data is outdated
- Reduce dependence on actual and labelled data:
  Mining and modelling drift

Key Idea

Substitution of **labelled data** by **knowledge of drift process**

Ex-post perspective:

- Study historic (labelled) data
- Identify and learn drift model: Explain change in $P(X, Y)$
- Verify model: Measure goodness of fit

Key Idea

Substitution of **labelled data** by **knowledge of drift process**

Ex-post perspective:

- Study historic (labelled) data
- Identify and learn drift model: Explain change in $P(X, Y)$
- Verify model: Measure goodness of fit

Ex-ante perspective:

- Update model on new (unlabelled) data: Extrapolate drift
- Classify using up-to-date model

Key Idea

Substitution of **labelled data** by **knowledge of drift process**

Ex-post perspective:

- Study historic (labelled) data
- Identify and learn drift model: Explain change in $P(X, Y)$
- Verify model: Measure goodness of fit

Ex-ante perspective:

- Update model on new (unlabelled) data: Extrapolate drift
- Classify using up-to-date model

Discussed Explicit Drift Models

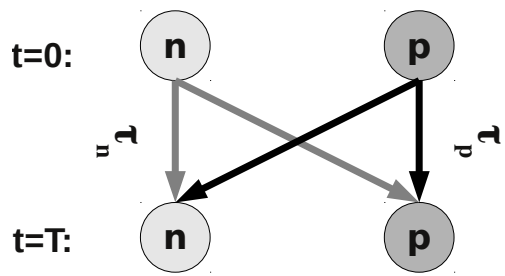
- Homogeneous Transitions Model
- Homogeneous Growth Model
- Drifting Subpopulations Model

Homogeneous Transitions Model Characteristics

- Static feature distribution, changing posterior distributions

Homogeneous Transitions Model Characteristics

- Static feature distribution, changing posterior distributions
- 2 states (negative, positive) for individuals
- Transitions between states:
  - Assumption: Markov property
  - State is observable (ex-post)
  - Intra-State transition probabilities: τ_p (pos.-pos.), τ_n (neg.-neg.)
  - Transition probabilities are not necessarily time-homogeneous



Homogeneous Transitions Model (2)

- Assumed change of the joint probability distribution:

$$P(x, y)_t = \tau_y P(x, y)_0 + (1 - \tau_{\bar{y}}) P(x, \bar{y})_0$$

Homogeneous Transitions Model (2)

- Assumed change of the joint probability distribution:

$$P(x, y)_t = \tau_y P(x, y)_0 + (1 - \tau_{\bar{y}}) P(x, \bar{y})_0$$

- Feature distribution is static:

$$P(x)_t = P(x)_0$$

- Class prior distribution changes:

$$P(y)_t = \tau_y P(y)_0 + (1 - \tau_{\bar{y}}) P(\bar{y})_0$$

- Posterior distributions change:

$$P(y|x)_t = \frac{\tau_y P(x, y)_0 + (1 - \tau_{\bar{y}}) P(x, \bar{y})_0}{P(x)_0}$$

Homogeneous Growth Model Characteristics

- Drifting feature and posterior distributions

Homogeneous Growth Model Characteristics

- Drifting feature and posterior distributions
- Two populations: negatives, positives
- Different (relative) growth factors: δ_p, δ_n

$$\delta_p = \frac{P(Y = p)_T}{P(Y = p)_0} \quad \text{and} \quad \delta_n = \frac{P(Y = n)_T}{P(Y = n)_0}$$

Homogeneous Growth Model (2)

- Assumed change of the joint probability distribution:

$$P(x, y)_t = \delta_y P(x, y)_0$$

Homogeneous Growth Model (2)

- Assumed change of the joint probability distribution:

$$P(x, y)_t = \delta_y P(x, y)_0$$

- Feature distribution changes:

$$P(x)_t = \delta_p P(x, p)_0 + \delta_n P(x, n)_0$$

- Class prior distribution changes:

$$P(y)_t = \delta_y P(y)_0$$

Homogeneous Growth Model (2)

- Assumed change of the joint probability distribution:

$$P(x, y)_t = \delta_y P(x, y)_0$$

- Feature distribution changes:

$$P(x)_t = \delta_p P(x, p)_0 + \delta_n P(x, n)_0$$

- Class prior distribution changes:

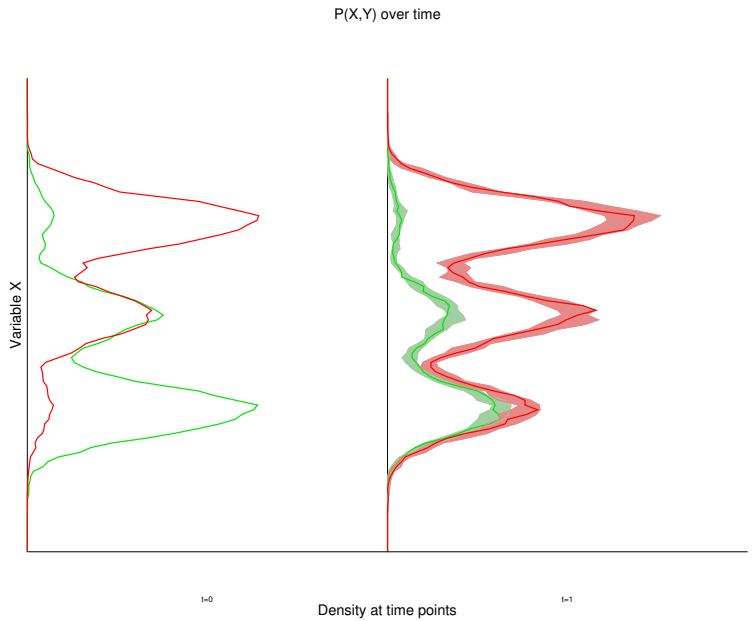
$$P(y)_t = \delta_y P(y)_0$$

- Posterior distributions change:

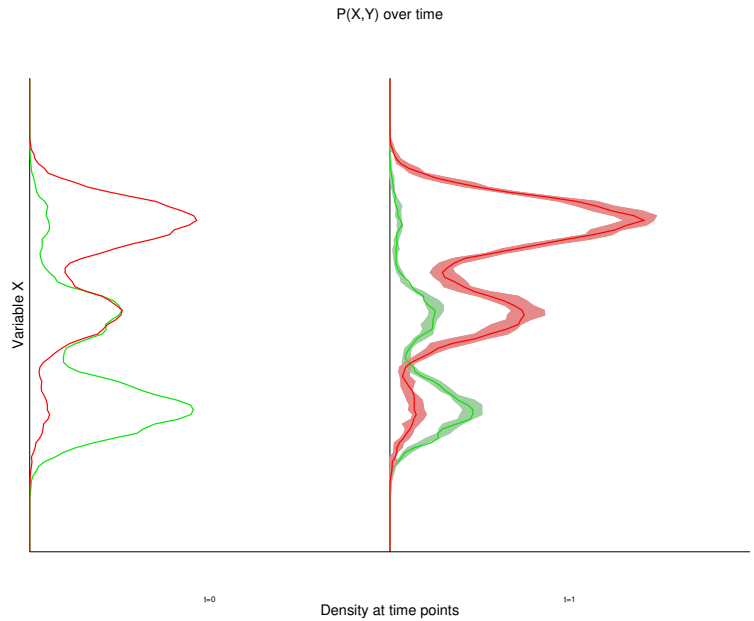
$$P(y|x)_t = \delta_y \xi_x P(y|x)_0 \qquad \text{with} \qquad \xi_x = \frac{P(x)_0}{P(x)_t}$$

Model Comparison

$P(X, Y)$ Drift in the Transition Model



$P(X, Y)$ Drift in the Growth Model



Local Drift

- Subpopulations are affected differently
- Two cases:
  - **Non-overlapping** feature distributions of subpopulations:
    Models from above can be applied locally

Local Drift

- Subpopulations are affected differently
- Two cases:
  - **Non-overlapping** feature distributions of subpopulations:
    Models from above can be applied locally
  - **Overlapping** or "evolving" feature distributions of subpopulations:
    Explicit mixture model of drifting subpopulations needed

    Approach:
    - Mixture decomposition (labelled data)
    - Mixture tracking (unlabelled data)

    Models and Algorithms:
    - Parametric case: Krempl, Hofer (2011)
    - Non-parametric case: Krempl (2011)

Assessing Drift in Data: Our Objectives and Approach

- Objectives
  - Is there drift in the data?
  - Can this drift be explained by a model?
  - Is the drift similar for all variables and years?

Assessing Drift in Data: Our Objectives and Approach

- Objectives
  - Is there drift in the data?
  - Can this drift be explained by a model?
  - Is the drift similar for all variables and years?
- Approach
  - Learn model on training sample
  - Evaluate goodness of fit on validation sample

Assessing Drift in Data: Our Objectives and Approach

- Objectives
  - Is there drift in the data?
  - Can this drift be explained by a model?
  - Is the drift similar for all variables and years?
- Approach
  - Learn model on training sample
  - Evaluate goodness of fit on validation sample
- Goodness of Fit Measures
  - Symmetric Kullback-Leibler divergence
  - Categorical: χ^2
  - Continuous: Kolmogorov - Smirnov, Anderson-Darling, Cramér-von-Mises

Experimental Setup

Dataset

- Random subsample of 58039 Danish companies
- Positive: Default in subsequent year,
  Negative: No default in any year
- Provided by the Danish Ministry of Taxation
- Labelled data from 2006-2010
- Discrete time variable (year of observation)
- Experiments *(for illustrational purpose only)*:
  Three groups of exemplary categorical variables:
  - A. Legal status
  - B. Region of operation
  - C. Branch of business

Variable A: Legal status, static model

Table: Drift Matrix of Posterior Drift in Variable A

	To,	Weighted Sym. KL Div.[a]			Sum of Sq. Diff.[b]		
	From	2007	2008	2009	2007	2008	2009
Static	2006	2.574	8.892	6.836	0.935	3.766	2.796
	2007	—	1.809	0.979	—	0.948	0.497
	2008	—	—	0.123	—	—	0.072

[a]: Weighted by $P(x)$, in 10^{-2} [b]: in 10^{-2}

Table: Drift Matrix: χ^2-Statistic on Variable A

	To,	χ^2 Statistic on $P(X, Y)$			χ^2 Statistic on $P(X)$		
	From	2007	2008	2009	2007	2008	2009
Static	2006	**134.00	**652.48	**462.74	**10.11	**55.57	**54.34
	2007	—	**99.39	**50.95	—	**16.09	**15.84
	2008	—	—	7.08	—	—	oo0.01

**: p-value $\leq$ 1%, *: p-value $\leq$ 10%, o: p-value $\geq$ 75%, oo: p-value $\geq$ 90%

Variable A: Legal status, static model

Table: Drift Matrix of Posterior Drift in Variable A

	To,	Weighted Sym. KL Div.[a]			Sum of Sq. Diff.[b]		
	From	2007	2008	2009	2007	2008	2009
Static	2006	2.574	8.892	6.836	0.935	3.766	2.796
	2007	—	1.809	0.979	—	0.948	0.497
	2008	—	—	0.123	—	—	0.072

[a]: Weighted by $P(x)$, in 10^{-2} [b]: in 10^{-2}

Table: Drift Matrix: χ^2-Statistic on Variable A

	To,	χ^2 Statistic on $P(X, Y)$			χ^2 Statistic on $P(X)$		
	From	2007	2008	2009	2007	2008	2009
Static	2006	**134.00	**652.48	**462.74	**10.11	**55.57	**54.34
	2007	—	**99.39	**50.95	—	**16.09	**15.84
	2008	—	—	7.08	—	—	oo0.01

**: p-value $\leq$ 1%, *: p-value $\leq$ 10%, o: p-value $\geq$ 75%, oo: p-value $\geq$ 90%

Variable A: Legal status, static model

Table: Drift Matrix of Posterior Drift in Variable A

	To,	Weighted Sym. KL Div.[a]			Sum of Sq. Diff.[b]		
	From	2007	2008	2009	2007	2008	2009
Static	2006	2.574	8.892	6.836	0.935	3.766	2.796
	2007	—	1.809	0.979	—	0.948	0.497
	2008	—	—	0.123	—	—	0.072

[a]: Weighted by $P(x)$, in 10^{-2} [b]: in 10^{-2}

Table: Drift Matrix: χ^2-Statistic on Variable A

	To,	χ^2 Statistic on $P(X, Y)$			χ^2 Statistic on $P(X)$		
	From	2007	2008	2009	2007	2008	2009
Static	2006	**134.00	**652.48	**462.74	**10.11	**55.57	**54.34
	2007	—	**99.39	**50.95	—	**16.09	**15.84
	2008	—	—	7.08	—	—	oo0.01

**: p-value $\leq$ 1%, *: p-value $\leq$ 10%, o: p-value $\geq$ 75%, oo: p-value $\geq$ 90%

Variable A: Legal status, static model

Table: Drift Matrix of Posterior Drift in Variable A

	To,	Weighted Sym. KL Div.[a]			Sum of Sq. Diff.[b]		
	From	2007	2008	2009	2007	2008	2009
Static	2006	2.574	8.892	6.836	0.935	3.766	2.796
	2007	—	1.809	0.979	—	0.948	0.497
	2008	—	—	0.123	—	—	0.072

[a]: Weighted by $P(x)$, in 10^{-2} [b]: in 10^{-2}

Table: Drift Matrix: χ^2-Statistic on Variable A

	To,	χ^2 Statistic on $P(X, Y)$			χ^2 Statistic on $P(X)$		
	From	2007	2008	2009	2007	2008	2009
Static	2006	**134.00	**652.48	**462.74	**10.11	**55.57	**54.34
	2007	—	**99.39	**50.95	—	**16.09	**15.84
	2008	—	—	7.08	—	—	oo0.01

**: p-value $\leq$ 1%, *: p-value $\leq$ 10%, o: p-value $\geq$ 75%, oo: p-value $\geq$ 90%

Variable A: Legal status

Table: Drift Matrix of Posterior Drift in Variable A

	To,	Weighted Sym. KL Div.[a]			Sum of Sq. Diff.[b]		
	From	2007	2008	2009	2007	2008	2009
Static	2006	2.574	8.892	6.836	0.935	3.766	2.796
	2007	—	1.809	0.979	—	0.948	0.497
	2008	—	—	0.123	—	—	0.072
Trans.M.	2006	0.786	2.077	1.689	0.233	0.883	0.666
	2007	—	0.595	0.341	—	0.231	0.122
	2008	—	—	0.003	—	—	0.002
Growth M.	2006	0.028	0.080	0.065	0.009	0.034	0.024
	2007	—	0.008	0.004	—	0.004	0.002
	2008	—	—	0.000	—	—	0.000

[a]: Weighted by $P(x)$, in 10^{-2} [b]: in 10^{-2}

Variable A: Legal status

Table: Drift Matrix of Posterior Drift in Variable A

	To,	Weighted Sym. KL Div.[a]			Sum of Sq. Diff.[b]		
	From	2007	2008	2009	2007	2008	2009
Static	2006	2.574	8.892	6.836	0.935	3.766	2.796
	2007	—	1.809	0.979	—	0.948	0.497
	2008	—	—	0.123	—	—	0.072
Trans.M.	2006	0.786	2.077	1.689	0.233	0.883	0.666
	2007	—	0.595	0.341	—	0.231	0.122
	2008	—	—	0.003	—	—	0.002
Growth M.	2006	0.028	0.080	0.065	0.009	0.034	0.024
	2007	—	0.008	0.004	—	0.004	0.002
	2008	—	—	0.000	—	—	0.000

[a]: Weighted by $P(x)$, in 10^{-2} [b]: in 10^{-2}

Variable A: Legal status

Table: Drift Matrix: χ^2-Statistic on Variable A

	To,	χ^2 Statistic on $P(X, Y)$			χ^2 Statistic on $P(X)$		
	From	2007	2008	2009	2007	2008	2009
Static	2006	**134.00	**652.48	**462.74	**10.11	**55.57	**54.34
	2007	—	**99.39	**50.95	—	**16.09	**15.84
	2008	—	—	7.08	—	—	^{oo}0.01
Trans.M.	2006	**83.71	**302.87	**233.17	**10.11	**55.57	**54.34
	2007	—	**116.52	**71.77	—	**16.09	**15.84
	2008	—	—	^{oo}0.22	—	—	^{oo}0.01
Growth M.	2006	^{oo}0.42	1.47	4.96	^{o}0.27	^{o}0.46	1.118
	2007	—	^{o}1.41	2.98	—	^{o}0.32	2.40
	2008	—	—	^{o}1.59	—	—	1.03

**: p-value $\leq$ 1%, *: p-value $\leq$ 10%, ^{o}: p-value $\geq$ 75%, ^{oo}: p-value $\geq$ 90%

Variable A: Legal status

Table: Drift Matrix: χ^2-Statistic on Variable A

	To,	χ^2 Statistic on $P(X, Y)$			χ^2 Statistic on $P(X)$		
	From	2007	2008	2009	2007	2008	2009
Static	2006	**134.00	**652.48	**462.74	**10.11	**55.57	**54.34
	2007	—	**99.39	**50.95	—	**16.09	**15.84
	2008	—	—	7.08	—	—	oo0.01
Trans.M.	2006	**83.71	**302.87	**233.17	**10.11	**55.57	**54.34
	2007	—	**116.52	**71.77	—	**16.09	**15.84
	2008	—	—	oo0.22	—	—	oo0.01
Growth M.	2006	oo0.42	1.47	4.96	o0.27	o0.46	1.118
	2007	—	o1.41	2.98	—	o0.32	2.40
	2008	—	—	o1.59	—	—	1.03

**: p-value $\leq$ 1%, *: p-value $\leq$ 10%, o: p-value $\geq$ 75%, oo: p-value $\geq$ 90%

Variable A: Legal status

Table: Drift Matrix: χ^2-Statistic on Variable A

	To,	χ^2 Statistic on $P(X, Y)$			χ^2 Statistic on $P(X)$		
	From	2007	2008	2009	2007	2008	2009
Static	2006	**134.00	**652.48	**462.74	**10.11	**55.57	**54.34
	2007	—	**99.39	**50.95	—	**16.09	**15.84
	2008	—	—	7.08	—	—	^{oo}0.01
Trans.M.	2006	**83.71	**302.87	**233.17	**10.11	**55.57	**54.34
	2007	—	**116.52	**71.77	—	**16.09	**15.84
	2008	—	—	^{oo}0.22	—	—	^{oo}0.01
Growth M.	2006	^{oo}0.42	1.47	4.96	^{o}0.27	^{o}0.46	1.118
	2007	—	^{o}1.41	2.98	—	^{o}0.32	2.40
	2008	—	—	^{o}1.59	—	—	1.03

**: p-value $\leq$ 1%, *: p-value $\leq$ 10%, ^{o}: p-value $\geq$ 75%, ^{oo}: p-value $\geq$ 90%

True vs. Predicted Class Prior Changes

True Prior Changes $\frac{P(p)_T}{P(p)_0}$

To / From	2007	2008	2009
2006	1.87	2.81	2.54
2007	—	1.50	1.36
2008	—	—	0.90

Predicted Prior Changes δ_p

To / From	2007	2008	2009
2006	1.84	2.82	2.52
2007	—	1.57	1.40
2008	—	—	0.88

True vs. Predicted Class Prior Changes

True Prior Changes $\frac{P(p)_T}{P(p)_0}$

To / From	2007	2008	2009
2006	1.87	2.81	2.54
2007	—	1.50	1.36
2008	—	—	0.90

Predicted Prior Changes δ_p

To / From	2007	2008	2009
2006	1.84	2.82	2.52
2007	—	1.57	1.40
2008	—	—	0.88

Detailed Results

	To, From	Predicted δ_p 2007	2008	2009
Var. A	2006	1.77	2.72	2.73
	2007	—	1.58	1.58
	2008	—	—	1.00
Var. B	2006	1.76	2.69	2.21
	2007	—	1.54	1.26
	2008	—	—	0.82
Var. C	2006	2.00	3.04	2.61
	2007	—	1.60	1.36
	2008	—	—	0.83

Summary and Conclusion

Summary

- Problem: Concurrence of *population drift* and *verification latency*
- Drift mining using explicit drift models

Summary and Conclusion

Summary

- Problem: Concurrence of *population drift* and *verification latency*
- Drift mining using explicit drift models
- Explicit drift models
  - Transition Model: 2 states, homogeneous transition probabilities τ_g, τ_b
  - Growth Model: Homogeneous growth factors δ_g, δ_b
  - Local: Drifting subpopulations, mixture model

Summary and Conclusion

Summary

- Problem: Concurrence of *population drift* and *verification latency*
- Drift mining using explicit drift models
- Explicit drift models
  - Transition Model: 2 states, homogeneous transition probabilities τ_g, τ_b
  - Growth Model: Homogeneous growth factors δ_g, δ_b
  - Local: Drifting subpopulations, mixture model
- Results on real-world company data:
  - Significant drift from 2006-2008, but not from 2008 to 2009
  - Extent of drift not increasing monotonically with time
  - Changes in data can be explained by a global growth model:

    No significant difference between $\hat{P}(X, Y)$ and $P(X, Y)$

Summary and Conclusion

Summary

- Problem: Concurrence of *population drift* and *verification latency*
- Drift mining using explicit drift models
- Explicit drift models
  - Transition Model: 2 states, homogeneous transition probabilities τ_g, τ_b
  - Growth Model: Homogeneous growth factors δ_g, δ_b
  - Local: Drifting subpopulations, mixture model
- Results on real-world company data:
  - Significant drift from 2006-2008, but not from 2008 to 2009
  - Extent of drift not increasing monotonically with time
  - Changes in data can be explained by a global growth model:

    No significant difference between $\hat{P}(X, Y)$ **and** $P(X, Y)$

Conclusion

Drift modelling and mining seems promising:

- **High goodness of fit**: Model can explain changes, no significant difference
- **Simplicity** of generative drift models
- **Prediction of drift of labelled data** by **changes in unlabelled data**

Bibliography

M. Böttcher, F. Höppner, and M. Spiliopoulou.

On exploiting the power of time in data mining.
ACM SIGKDD Explorations Newsletter, 10(2):3–11, 2008.

J. N. Crook, L. C. Thomas, and R. Hamilton.

The degradation of the scorecard over the business cycle.
In L. C. Thomas, D. B. Edelman, and J. N. Crook, editors, *Readings in Credit Scoring*, pages 161–175. Oxford University Press, 2004.

C. Hoyland.

Recessions and recession scoring.
In M. Bailey, editor, *Credit Scoring - The Principles and Practicalities*, pages 133–137. White Cox Publishung, 2 edition, 2004.

M. G. Kelly, D. J. Hand, and N. M. Adams.

The impact of changing populations on classifier performance.
In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 367–371, 1999.

L. I. Kuncheva.

Classifier ensembles for detecting concept change in streaming data: Overview and perspectives.
In O. Okun and G. Valentini, editors, *Proceedings of the second workshop on supervised and unsupervised ensemble methods and their applications (SUEMA2008)*, volume 245 of *Studies in Computational Intelligence*. Springer, 2008.

A. Lucas.

Updating scorecards: Removing the mystique.
In L. C. Thomas, D. B. Edelman, and J. N. Crook, editors, *Readings in Credit Scoring*, pages 93–110. Oxford University Press, 2004.

G. Marrs, R. Hickey, and M. Black.

The impact of latency on online classification learning with concept drift.
In Y. Bi and M.-A. Williams, editors, *Knowledge Science, Engineering and Management*, volume 6291 of *Lecture Notes in Computer Science*, pages 459–469. Springer, 2010.

M. Spiliopoulou, I. Ntoutsi, Y. Theodoridis, and R. Schult.

The monic framework for cluster transition detection.
In *Proc. of the 5th Hellenic Data Management Symposium*, 2006.

S. Trueck and S. T. Rachev.

Rating Based Modeling of Credit Risk.
Elsevier, 2009.

Homogeneous Growth Model: Ex-ante parameter estimation

Minimisation Problem

Given $P(X)_1$ and $P(X,Y)_0$, determine δ_p and δ_n such that

$$SSE = \sum_x \left(\hat{P}(x)_1 - P(x)_1\right)^2 P(x)_1 \rightarrow \min$$

(Ex-ante) Parameter Estimation

$$\delta_p = \frac{\sum_x w_x P(x)_1^2 - \sum_x w_x P(x|n)_0 P(x)_1}{\sum_x w_x^2 P(x)_1}$$

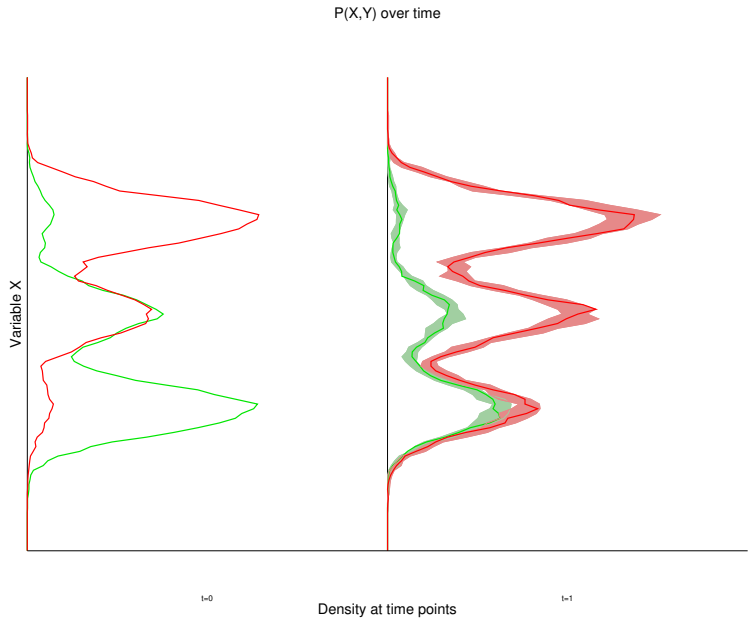
$$\delta_n = \frac{1 - \delta_p P(p)_0}{P(n)_0}$$

with

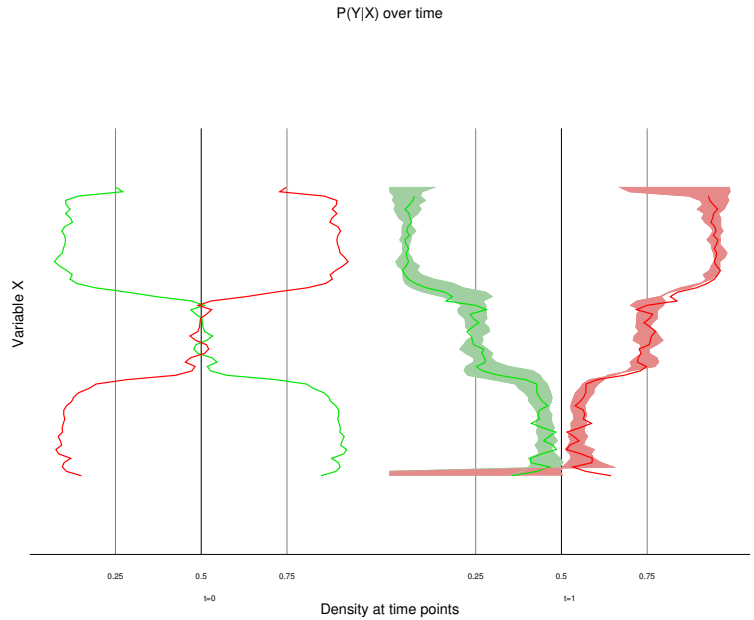
$$w_x = P(x,p)_0 - P(x,n)_0 \frac{P(p)_0}{P(n)_0}$$

Homogeneous Transitions Model (3)

$P(X, Y)$ over time

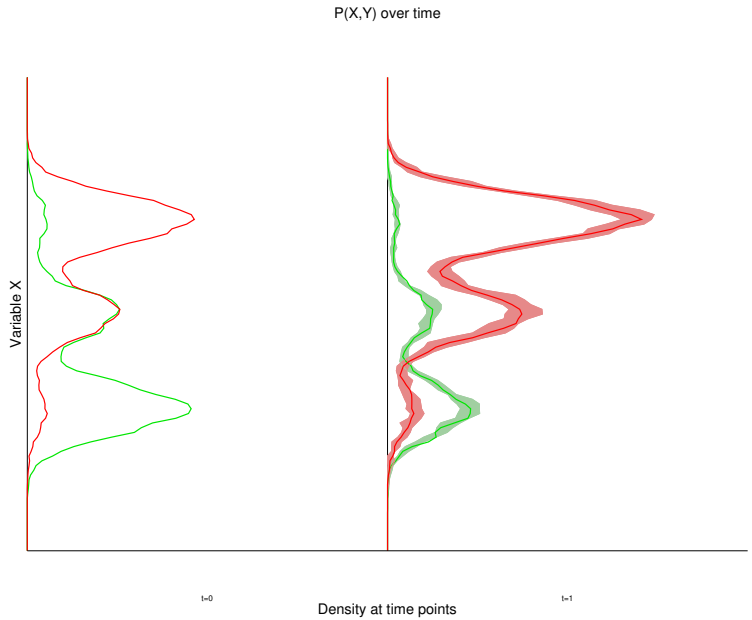


$P(Y|X)$ over time

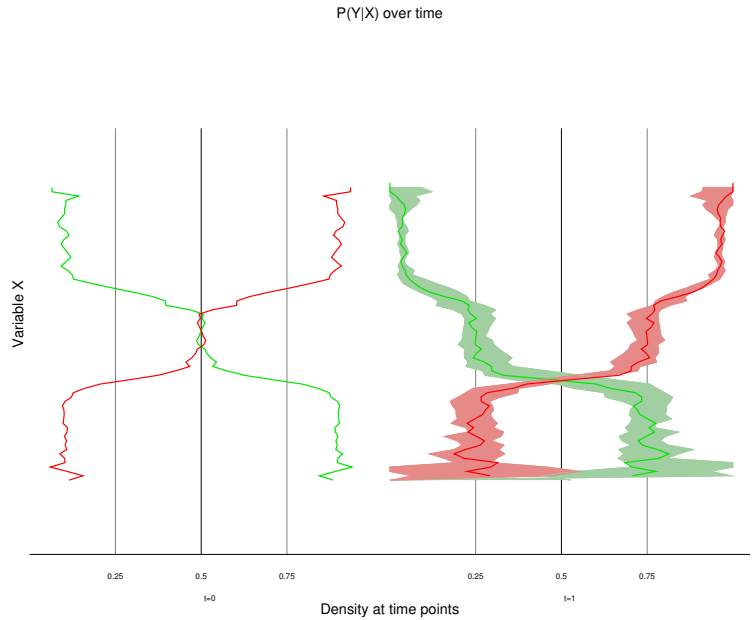


Homogeneous Growth Model (3)

$P(X, Y)$ over time

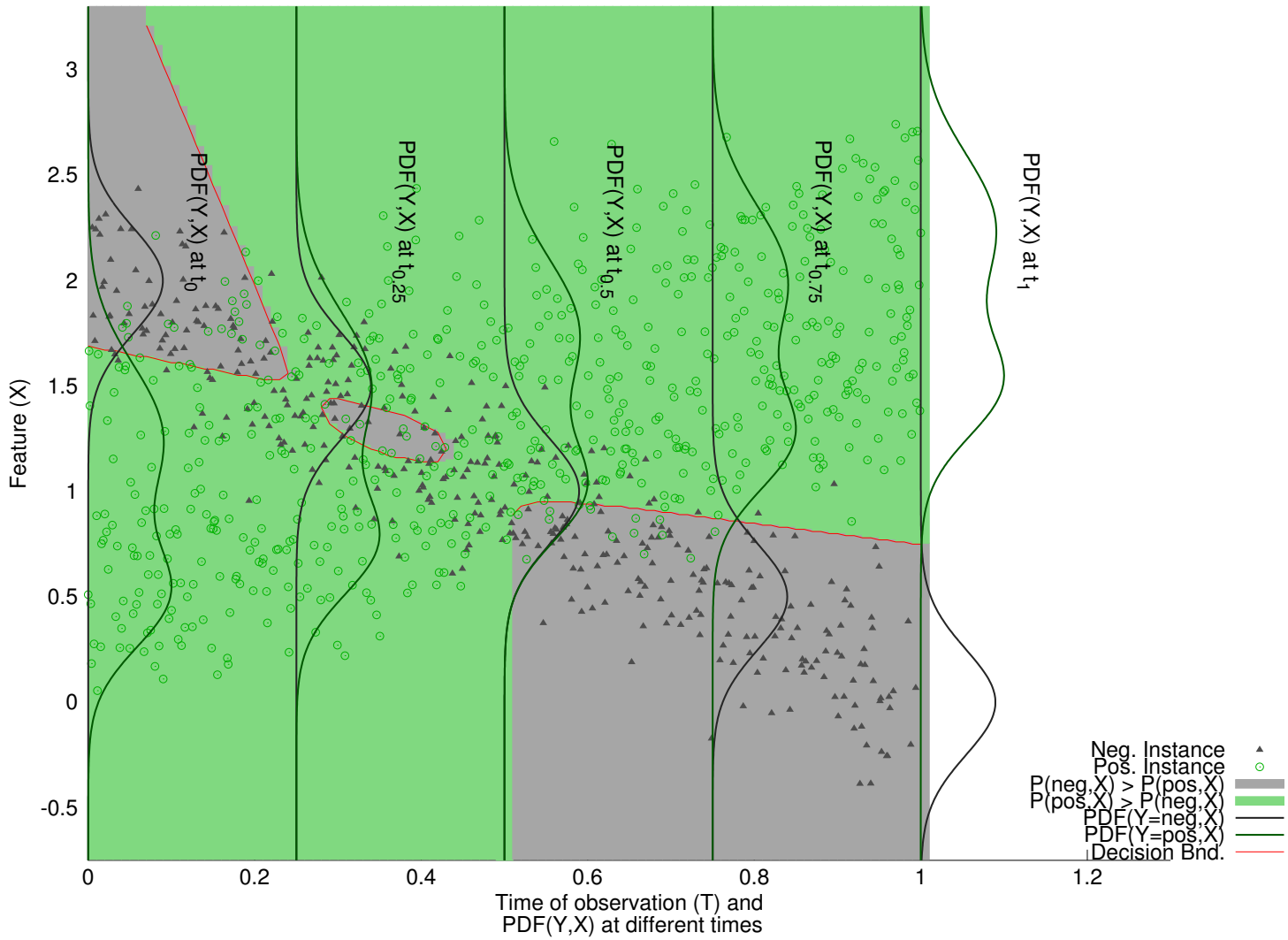


$P(Y|X)$ over time



Local Drift: Drifting Subpopulations

Illustration



Goodness of Fit: Notation

Notation:

- $\hat{P}(Y|X)_T$ Estimated posterior at time $t = T$,
- $P(Y|X)_T$ Observed (empirical) posterior at time $t = T$,
- $P(X)_T$ Observed (empirical) feature distribution at time $t = T$.

Symmetric, weighted Kullback-Leibler Div.

Divergence of the posteriors weighted by the feature distribution

$$SymKL_{weighted}(\hat{P}(Y|X)_T, P(Y|X)_T) =$$
$$\sum_x P(x)_T \left(KL(\hat{P}(Y|x)_T \parallel P(Y|x)_T) + KL(P(Y|x)_T \parallel \hat{P}(Y|x)_T) \right)$$

Sum of squared posterior differences

$$SSE(\hat{P}(Y|X)_T, P(Y|X)_T) = \sum_x \left(\hat{P}(Y|x) - P(Y|x) \right)^2$$

Drift in Class Prior

Table: Change in Class Prior over Time

Year	2006	2007	2008	2009
$P(n)$	95.46%	91.51%	87.26%	88.48%
$P(p)$	4.54%	8.49%	12.74%	11.52%

Variable B: Region of operation

Table: Drift Matrix of Posterior Drift in Variable B

	To,	Weighted Sym. KL Div.[a]			Sum of Sq. Diff.[b]		
	From	2007	2008	2009	2007	2008	2009
Static	2006	2.574	8.892	6.836	0.935	3.766	2.796
	2007	—	1.809	0.979	—	0.948	0.497
	2008	—	—	0.123	—	—	0.072
Trans.M.	2006	0.786	2.077	1.689	0.233	0.883	0.666
	2007	—	0.595	0.341	—	0.231	0.122
	2008	—	—	0.003	—	—	0.002
Growth M.	2006	0.028	0.080	0.065	0.009	0.034	0.024
	2007	—	0.008	0.004	—	0.004	0.002
	2008	—	—	0.000	—	—	0.000

[a]: Weighted by $P(x)$, in 10^{-2} [b]: in 10^{-2}

Variable B: Region of operation

Table: Drift Matrix: χ^2-Statistic on Variable B

	To,	χ^2 Statistic on $P(X, Y)$			χ^2 Statistic on $P(X)$		
	From	2007	2008	2009	2007	2008	2009
Static	2006	**139.95	**660.93	**510.64	*5.86	**32.78	**16.21
	2007	—	**97.73	**54.27	—	**9.57	2.13
	2008	—	—	**11.90	—	—	2.44
Trans.M.	2006	**51.06	**198.31	**100.17	*5.86	**32.78	**16.21
	2007	—	**62.87	**13.48	—	**9.57	2.13
	2008	—	—	*7.787	—	—	2.44
Growth M.	2006	3.09	3.19	**19.58	0.11	0.12	1.07
	2007	—	^{oo}0.31	4.45	—	^{o}0.06	0.27
	2008	—	—	*7.12	—	—	0.57

**: p-value $\leq$ 1%, *: p-value $\leq$ 10%, ^{o}: p-value $\geq$ 75%, ^{oo}: p-value $\geq$ 90%

Variable C: Branch of business

Table: Drift Matrix of Posterior Drift in Variable C

	To,	Weighted Sym. KL Div.[a]			Sum of Sq. Diff.[b]		
	From	2007	2008	2009	2007	2008	2009
Static	2006	2.574	8.892	6.836	0.935	3.766	2.796
	2007	—	1.809	0.979	—	0.948	0.497
	2008	—	—	0.123	—	—	0.072
Trans.M.	2006	0.786	2.077	1.689	0.233	0.883	0.666
	2007	—	0.595	0.341	—	0.231	0.122
	2008	—	—	0.003	—	—	0.002
Growth M.	2006	0.028	0.080	0.065	0.009	0.034	0.024
	2007	—	0.008	0.004	—	0.004	0.002
	2008	—	—	0.000	—	—	0.000

[a]: Weighted by $P(x)$, in 10^{-2} [b]: in 10^{-2}

Variable C: Branch of business

Table: Drift Matrix: χ^2-Statistic on Variable C

	To,	χ^2 Statistic on $P(X, Y)$			χ^2 Statistic on $P(X)$		
	From	2007	2008	2009	2007	2008	2009
Static	2006	**135.71	**659.66	**464.89	*4.76	**22.57	**13.83
	2007	—	**97.80	**48.14	—	*5.81	2.01
	2008	—	—	5.68	—	—	0.90
Trans.M.	2006	**31.24	**102.48	**76.98	*4.76	**22.57	**13.83
	2007	—	**28.71	**14.95	—	*5.81	2.01
	2008	—	—	^{o}1.03	—	—	0.90
Growth M.	2006	^{o}1.16	3.75	2.75	^{o}0.08	0.31	^{o}0.05
	2007	—	^{oo}0.48	^{oo}0.17	—	0.15	^{oo}0.00
	2008	—	—	^{oo}0.17	—	—	0.16

**: p-value $\leq$ 1%, *: p-value $\leq$ 10%, ^{o}: p-value $\geq$ 75%, ^{oo}: p-value $\geq$ 90%