Master Thesis: An Evaluation of Calibration Methods for Data Mining Models in Simulation Problems

Antonio Bella Sanjuán

Supervisors: César Ferri Ramírez
José Hernández Orallo
Maria José Ramírez Quintana

DEPARTAMENT DE SISTEMES INFORMÀTICS I COMPUTACIÓ

Valencia, 18 December 2008

Outline

- Introduction
- Calibration of Machine Learning Models
- Simulation in Multi-Decision Data Mining Problems
- Contributions
- Conclusion
- Future Work

Outline

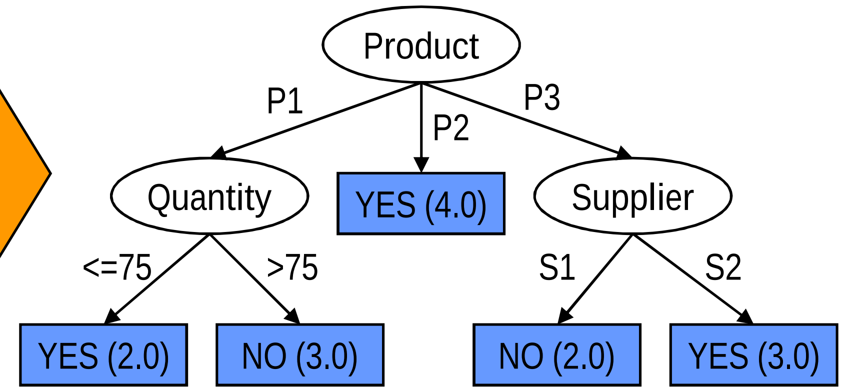
- Introduction
- Calibration of Machine Learning Models
- Simulation in Multi-Decision Data Mining Problems
- Contributions
- Conclusion
- Future Work

Introduction (I)

Training Data

Supplier	Product	Quantity	Price	Delivered on time?
S1	P1	85	85	NO
S2	P1	90	80	NO
S1	P2	86	83	YES
S1	P3	96	70	YES
S1	P3	80	68	YES
S2	P3	70	65	NO
S2	P2	65	64	YES
S1	P1	95	72	NO
S1	P1	70	69	YES
S1	P3	80	75	YES
S2	P1	70	75	YES
S2	P2	90	72	YES
S1	P2	75	81	YES
S2	P3	91	71	NO

Data Mining Model



Decision Problem

Supplier	Product	Quantity	Price
S1	P1	70	70
S2	P1	80	75

Introduction (II)

The Fastest Supplier DM Model

Supplier	...	Prob.
S1	...	0.9
S2	...	0.7
S3	...	0.5
S4	...	0.2

*

The Cheapest Supplier DM Model

Supplier	...	Prob.
S4	...	0.8
S3	...	0.6
S2	...	0.4
S1	...	0.3

Fast and Cheap Supplier

Supplier	Fast & Cheap
S3	0.30
S2	0.28
S1	0.27
S4	0.16

- Problem: The best local decisions do not make the best local result
- Solution: Combine local models and then, use simulation to obtain a good global result

Introduction (III)

The Fastest Supplier DM Model

Supplier	...	Prob. 1
S1	...	0.9
S2	...	0.7
S3	...	0.5
S4	...	0.2

*

The Cheapest Supplier DM Model

Supplier	...	Prob.
S4	...	0.8
S3	...	0.6
S2	...	0.4
S1	...	0.3

Fast and Cheap Supplier

Supplier	Fast & Cheap 1	Fast & Cheap 2
S3	0.30	0,18
S2	0.28	0,20
S1	0.27	0,21
S4	0.16	0

- Problem: Combine several non-realistic probabilistic models can make the overall model diverge
- Solution: Calibrate the estimated probabilities

Outline

- Introduction
- Calibration of Machine Learning Models
- Simulation in Multi-Decision Data Mining Problems
- Contributions
- Conclusion
- Future Work

Calibration of Machine Learning Models (I)

- Estimated probability: 0.9
- Actual frequency: 50%
- **UNCALIBRATED!!!** Estimation too optimistic

- Estimated probability: 0.5
- Actual frequency: 90%
- **UNCALIBRATED!!!** Estimation too pessimistic

- Estimated probability: 0.9
- Actual frequency: 88%
- **CALIBRATED!!!** Realistic estimation

Calibration of Machine Learning Models (II)

- State of the art
- Established a **taxonomy** (calibration techniques and measures)
- **Clarification of the calibration concept**
- New **multivariate calibration method** versus univariate classical calibration methods.
- It is a **multi-class calibration method** versus binary-class classical calibration methods
- Experimental evaluation:
  - 2 calibration measures: CalBin and MSE
  - 4 calibration methods: Binning Averaging, Isotonic Regression (PAV), Platt's Method and **Similarity-Binning Averaging**
  - 2 baseline methods: Base (without calibration) and 10-NN

Calibration of Machine Learning Models (III)

- Experimental Results: Column vs. Row Nemenyi Test: V win, = tie, X loss

CalBin Measure

10-NN	Bin	PAV	Platt	SB	
V	X	=	X	V	Base
	X	X	X	V	10-NN
		=	V	V	Bin
			X	V	PAV
				V	Platt

MSE Measure

10-NN	Bin	PAV	Platt	SB	
V	X	=	V	V	Base
	X	V	X	V	10-NN
		V	V	V	Bin
			=	V	PAV
				V	Platt

Outline

- Introduction
- Calibration of Machine Learning Models
- Simulation in Multi-Decision Data Mining Problems
- Contributions
- Conclusion
- Future Work

Simulation in Multi-Decision DM Problems (I)

Campaign Design with *N* Products

- Several products to be offered to a house list of customers, with constraints (stock limitations, costs,...) → One DM model for each product
- Use simulation with Petri nets to obtain better cutoffs for the data mining models (fulfilling the constraints)
- Experimental results:
  - Single baseline method (without simulation)
  - Joint simulation method

Simulation in Multi-Decision DM Problems (II)

Similarity-Binning Calibration Applied to Campaign Design with *N* Products

- Experimental Results: Wilcoxon Signed-Ranks Test: = no significant differences, > or < significant differences
- Single baseline method vs. Joint simulation method
- Non-calibrated vs. Calibrated (Similarity-Binning Averaging)

	2 products			3 products			4 products		
	Benefit non-calibrated models		Benefit calibrated models	Benefit non-calibrated models		Benefit calibrated models	Benefit non-calibrated models		Benefit calibrated models
Single	4181	=	5230	4881	=	8150	-5986	<	-5229
Joint	10074	>	9246	22562	=	21300	8445	<	9112

Outline

- Introduction
- Calibration of Machine Learning Models
- Simulation in Multi-Decision Data Mining Problems
- Contributions
- Conclusion
- Future Work

Contributions

- *Joint Cutoff Probabilistic Estimation Using Simulation: A Mailing Campaign Application*. 8th International Conference on Intelligent Data Engineering and Automated Learning (**IDEAL 2007**). **Springer Verlag. LNCS 4881**
- *Calibration of Machine Learning Models.* Chapter of the *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods and Techniques.* **IGI Global.**
- *Similarity-Binning Averaging: A Generalisation of Binning Calibration.* ***(submitted to KDD 2009)***

Outline

- Introduction
- Calibration of Machine Learning Models
- Simulation in Multi-Decision Data Mining Problems
- Contributions
- Conclusion
- Future Work

Conclusion

- Simulation to combine local data mining models and obtain good overall results
- Taxonomy of calibration measures and methods
- New multi-class calibration method
- Good performance

Outline

- Introduction
- Calibration of Machine Learning Models
- Simulation in Multi-Decision Data Mining Problems
- Contributions
- Conclusion
- Future Work

Future Work

- New scenarios
- Intelligent agent negotiation between seller and buyer
- Compare our calibration method with existing multi-class calibration approaches

Thanks for your attention!