Markov-Chain based Credit Control for Subscribers to Mobile Communication Services

Falk Wagner

Credit Scoring and Credit Control XII

Edinburgh; Aug 25, 2011

Motivation of the research

- Market saturation intensifies competition among operators and decreases profit margins.
- To encounter this development and grow profits, operators may …
  - Increase revenues from subscribers, → Profit Scoring
  - Optimize acquisition costs, → Risk- / profit-based alternative offers
  - Cut costs and → Reducing bad debt
  - Increase subscriptions' lifetime. → Optimizing credit control
- Credit control can be a powerful instrument to prevent default on payment due to overspending (credit churn).
- Based on sophisticated Credit Limits, subscriber spendings can be put under thorough control (e.g. by deactivating high cost, but low margin premium rate services once a limit has been hit).
- This research looks into optimizing the calculation of credit limits by applying Markov Chains models to account data of a Mobile Communication Service provider.

- Markov Chains are a modeling technique from statistics to formalize discrete-time stochastic process.
- They are composed of a sequence of states each describing the status of a system at a certain point in time.

	Current + immediate preceding period	Current + n preceding periods
1) Path dependence	*First-order* Markov Chain	*Higher-order* Markov Chain (Second, Third etc.)

	Time independent	Time dependent
2) Time dependence	*Stationary* Markov Chain	*Non-Stationary* Markov Chain

- A *Transition Matrix* describes the probability of the system to move to a certain state given a history of states visited previously.

Markov Chain-based modelling of Credit Risk

Month I Status (May)
Current
Arrears 1
Arrears 2
Arrears 3
Arrears 4
Total

- Markov Chains were introduced to credit risk modelling by (Cyert, Davidson et al. 1962).
- (Mehta 1970) and (Liebman 1972) provide early examples of Markov Chain-based optimization of credit policies.
- (Frydman, Kallberg et al. 1985) introduced *Population Heterogeneity* of accounts by distinguishing between objects making moves across the Transition Matrix – so called *Movers* – and objects never leaving their initial state – so called *Stayers*.

- Credit churn is driven by overspending. We‛re thus interested to understand which changes in an account‛s spending behaviour are associated to high risk of default.

- Analytically, three model components need to worked on:

  1. Relevant features of the account as of period t need to be identified and translated into a *state(*t*)* identifier.
  2. Relevant features of the account as of period t+1 need to be identified and translated into a *state(*t+1*)* identifier.

  - The *state(*t+1*)* identifier should be based on account-related information as available to identifier *state(*t*)* but on spending data as of period t+1.
  - *State(*t+1*)* thus represents the account as of period t but under changed spending behaviour.
  - Plotting of combinations of *state(*t*)* and *state(*t+1*)* creates a *Transition matrix*.

  3. As we‛re interested in how risk changes as spendings vary, the Transition matrix‛ structure will be used to calculate the risk associated to individual moves (*Risk Matrix*).

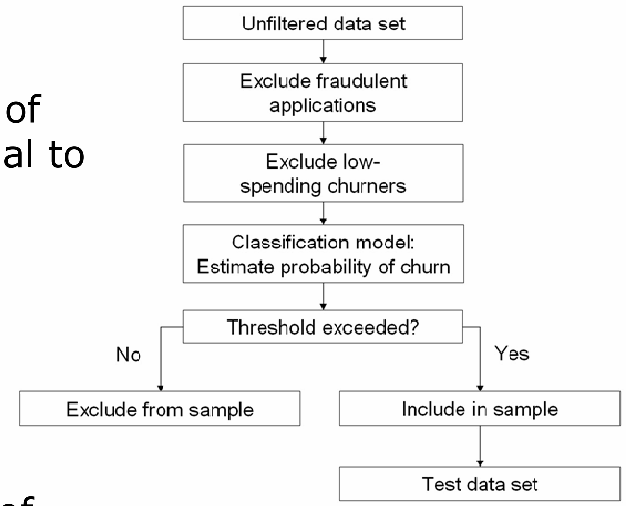
- The research uses data of 80.000 subscribers to mobile communication services. For each customer invoice data and aggregated usage patterns for up to 24 billing cycles were available.
- Due to the large amount of data at hand and it being blurred by side effects, the data needed to be pre-processed.

Pre-processing of the data set:

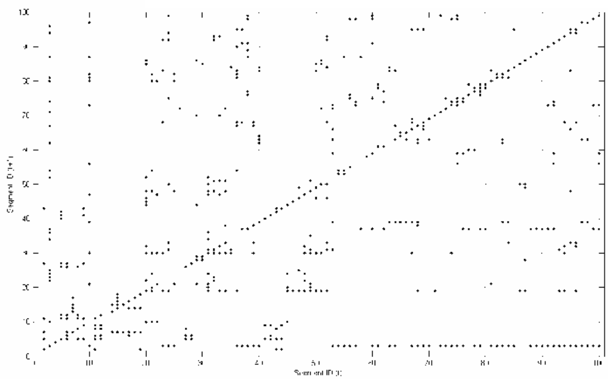
1. Default on payment of fraudulent subscribers can't be prevented by Credit control („no intention to pay").

2. Default on payment on amounts at / below the average of previous billing cycles is likely caused by reasons external to the customer relationship.

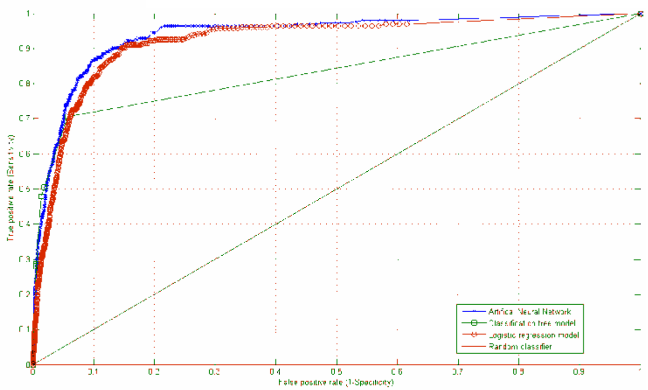
3. *Stayer* accounts may be eliminated:
   - No sufficient examples of „default" for this group.
   - As invoice amounts are very steady, Credit Limits will likely never apply.
   - A *Behaviour Score* may support the identification of such low-risk segment.



- A Behaviour Score was to model the probability of an account to default in the upcoming billing cycle.
- Accounts with a probability of < 1% were to be removed from the sample.
- To identify the best-suited Data Mining algorithm for the problem, a benchmark was performed:



Algorithm	AUC on Training Sample	AUC on Test Sample	AUC on Hold-out Sample	Accuracy at cut-off	False positive rate at cut-off
Logistic regression	0.92809	0.91784	0.92084	0.88333	0.14902
Decision tree	0.83097	0.79659	0.83388	0.94182	0.29412
ANN	0.94249	0.92549	0.93847	0.87268	0.10588

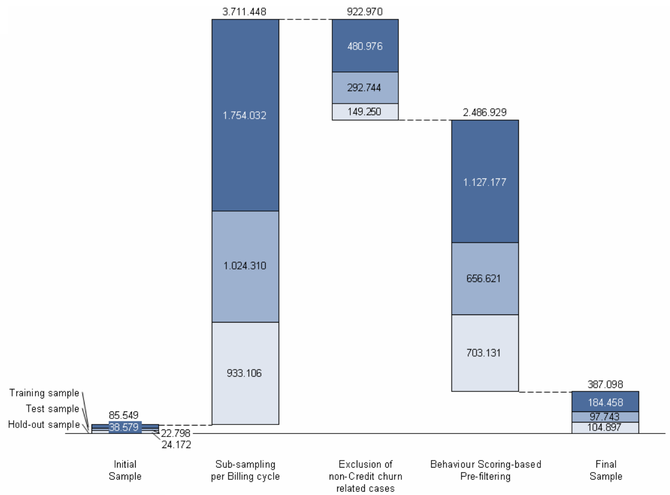


- ANN outperformed LR and DT in terms of AUC.
- ANN and DT ROC curves intersected in a small area of the performance space.
- Better overall performance and a lower False-positive rate at the cut-off made us choose ANN.

Behaviour Scoring-based sub-sampling

- The Behaviour Score allowed to dramatically reduce the sample size without losing many examples of credit churn.
- The sample was used to develop and compare three Markov Chain-variants:

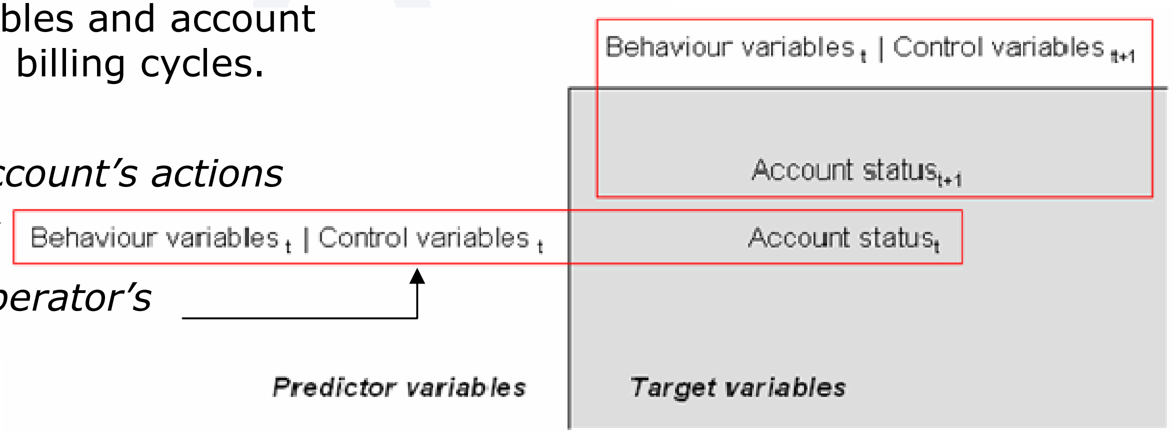
Markov Chain variant	Predictors
First-order stationary	Data of periods t and t+1
First-order non-stationary	Data of periods t and t+1 + Identifier of period t
Second-order non-stationary	Data of periods t-1, t and t+1 + Identifier of period t



- The samples where compiled based on Behaviour and Control variables and account status as of the idendividual billing cycles.

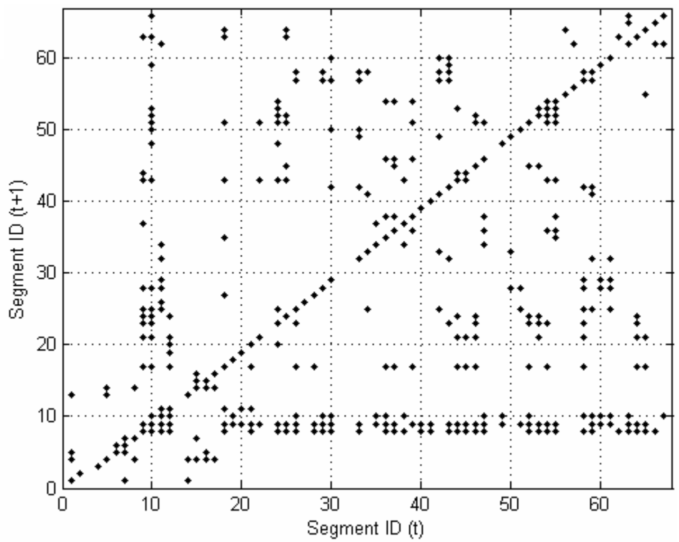
1) Variables describing the account's actions

2) Variables describing the operator's actions towards the account



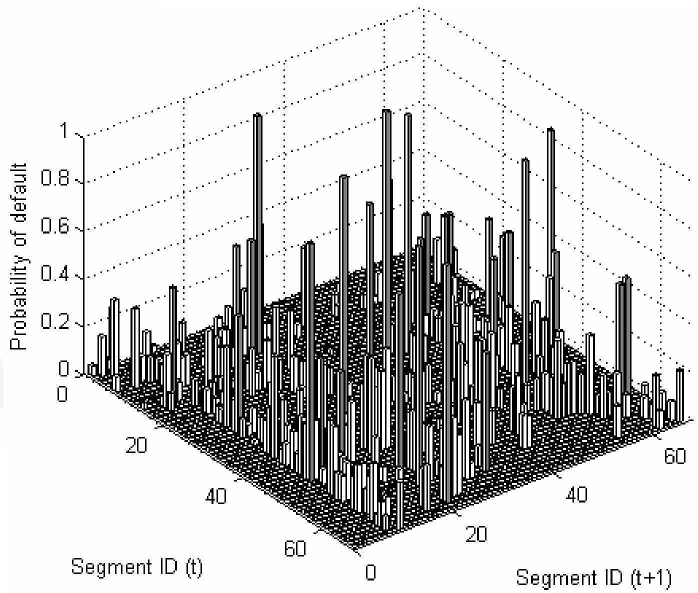
Identifying *State* pattern:

- A Decision Tree was used to group *Behaviour* and *Control variables*.
- The model used “Probability of default in t+1” as target.
- The DT nodes individual records were assigned to, were used as representation of the account *state* in that period.



Calculation of the *Risk Matrix*:

- The model was run for each billing cycle.
- *Nodes*(t) and *Nodes*(t+1) were stored.
- Based on *Nodes*(t) and *Nodes*(t+1) the *Transition matrix* was populated.
- Measuring the probability of default per Node combination allowed to calculate the expected probability of default.

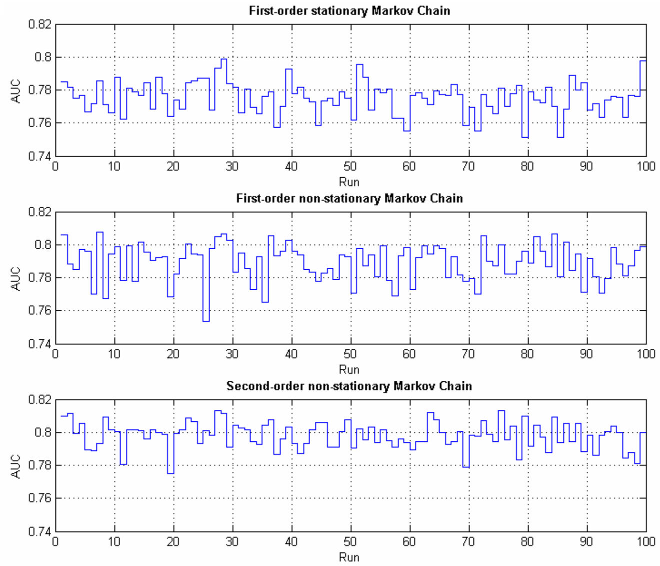


mobile business

Benchmarking the Markov Chain variants

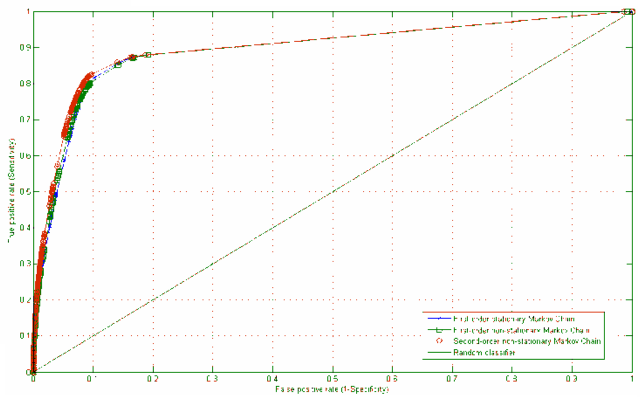
- To identify the best performing Markov Chain variant, an AUC benchmark was performed:

	Mean AUROC	Maximum AUROC	Minimum AUROC
First order stationary Markov Chain	0.7749	0.7988	0.7511
First order non-stationary Markov Chain	0.7892	0.8075	0.7537
Second order non-stationary Markov Chain	0.7980	0.8131	0.7748

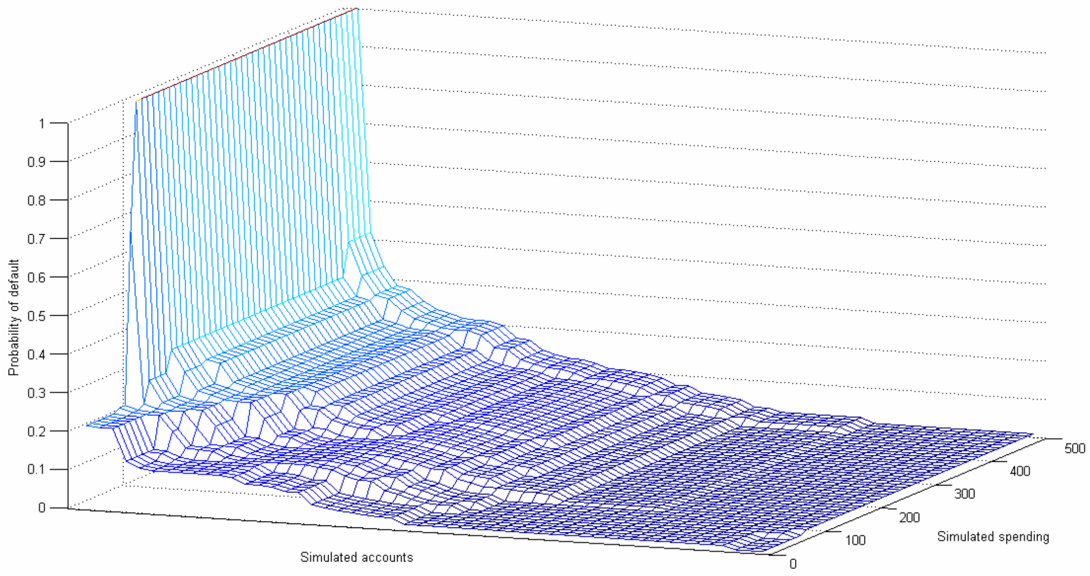


- It could be found that the Second-order non-stationary Markov Chain provided the best performance.

- The only slim performance margin suggests that *Account history carried* and *Time dependence* are of only minor influence.



Application of the model



- The *Risk Matrix* allows to estimate the probability of default as a function of varied spending levels.
- By simulating spendings, the risk of indivual accounts to default at selected spending levels can be estimated.

- The simulation data allows to calculate an subscriber-individual Credit Limit that is in line with the operator's risk acceptance policy.
- By applying such Credit Limit, the risk of Credit churn can be reduced.
- Economically, the increased subscription lifetime increases the operator's rentability of the accounts handeled according to the optimized strategy.

Thank you.

falk.wagner@m-chair.net

References

- Cyert, R. M., H. J. Davidson, et al. (1962). "Estimation of the Allowance for Doubtful Accounts by Markov Chains." Management Science 8(3): 287-303.
- Frydman, H., J. G. Kallberg, et al. (1985). "Testing the Adequacy of Markov Chain and Mover-Stayer Models as Representations of Credit Behavior." Operations Research 33(6).
- Liebman, L. H. (1972). "A Markov Decision Model for Selecting Optimal Credit Control Policies." Management Science 18(10): B519-B525.
- Mehta, D. (1970). "Optimal Credit Policy Selection: A Dynamic Approach." Journal of Financial and Quantitative Analysis 5: 421-444.