Bayesian Networks in Credit Rating

Samuel Gerssen

March 12, 2004

Prof. Dr. H. Koppelaar
Drs. Dr. L.J.M. Rothkrantz
Ir. W. de Jong
Dr. K. van der Meer
Dr. D. Fokkema (ABN AMRO)
Dr. Ir. M.J. Druzdzel (University of Pittsburgh)
Bayesian Networks in Credit Rating
Contents

- Credit Rating
- Bayesian Networks
- Assignment
- Bayesian Networks Learning Research
- Bayesian Networks in Credit Rating
- Conclusions
Credit Rating - Definition

Credit Rating = Assessment of risk on credit portfolios

Total Outstanding Credit (Credit Portfolio)

Loss

No Loss

Economic Capital

0% EL 100%
Credit Rating – Expected Loss per client

Loss per client:

- Default?
  - N: No Loss
  - Y: Loss = EAD ($) x LGD (%)

Expected Loss per client = PD (EAD x LGD)
Credit Rating – Expected Loss

\[ EL = PD \times (EAD \times LGD) \]

PD = Probability of Default
EAD = Exposure At Default
LGD = Loss Given Default
Credit Rating – Probability of Default

Probability of Default (PD)

= the probability that a client can not meet its repayment obligations between now and 1 year
Credit Rating – PD Model

PD calculation for each client

<table>
<thead>
<tr>
<th>Client</th>
<th>Variable 1</th>
<th>Variable 2</th>
<th>...</th>
<th>Variable N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

 PD
Credit Rating – Modeling steps

Construction of data set:
- Select portfolio information from previous year
- Add default history from last 12 months

Modeling technique: Binary logistic regression

Scoring: 5-fold cross validation
Credit Rating - Binary Logistic Regression

PD depends on variables \((x_1 .. x_n)\) and parameters \((\beta_0 .. \beta_n)\)

\[
\ln\left(\frac{PD}{1-PD}\right) = \beta_0 + \beta_1 x_1 + .. + \beta_n x_n
\]

Calculation of \((\beta_0 .. \beta_n)\) using maximum likelihood
A Bayesian Network (BN) is a probabilistic model of variables and their causal relations.

A BN is a directed acyclic graph where:

Variables are nodes.

Causal relations are vertices.
Conditional Probability Tables (CPT)

\[
\begin{align*}
P(A) &= 0.02 \\
P(B) &= 0.3 \\
P(C | AB) &= 0.95 \\
P(C | A \neg B) &= 0.8 \\
P(C | \neg AB) &= 0.1 \\
P(C | \neg A \neg B) &= 0.01
\end{align*}
\]
Bayesian Networks – Inference

- Bayes’ Theorem

\[ p(a \mid b) = \frac{p(b \mid a)p(a)}{p(b)} \]

- Expansion rule

\[ p(a) = p(ab) + p(ab\bar{b}) = p(a \mid b)p(b) + p(a \mid \bar{b})p(\bar{b}) \]
\[ p(a) = p(abc) + p(ab\bar{c}) + p(a\bar{b}c) + p(a\bar{b}\bar{c}) \]
Bayesian Networks - Learning

\[ \theta = \text{probability of variable } X \]

\[ p(\theta) = \text{probability distribution of } \theta \]

= belief about \( \theta \)

Example of \( p(\theta) \)
Bayesian Networks – Belief Updating

Observation: $O_i$

The belief about $\theta$ is now updated:

$$p(\theta | O_i) = \begin{cases} 
    c \cdot \theta \cdot p(\theta) & \text{if } O_i : X = 1 \\
    c \cdot (1-\theta) \cdot p(\theta) & \text{if } O_i : X = 0
\end{cases}$$

$c = \text{normalization coefficient}$
Observation $O_i : X=1$

Belief about $\theta$ is updated

Updated distribution $p(\theta|O_i) = p(\theta) \cdot \theta$
Bayesian Networks – Beta distribution

Beta distribution with parameters $\alpha$ and $\beta$:

$$p(\theta \mid \alpha \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$\alpha$ = the number of observations $X_i = 1$

$\beta$ = the number of observations $X_i = 0$

Property: after updating again Beta distribution
Bayesian Networks - Structures

Applied network structures in this research:

- CPT model
- Naive Bayes model
- Noisy-MAX model
Bayesian Networks – CPT model

Size of CPT = $(s_p)^p$

$s_p = \text{number of parent states}$

$p = \text{number of parents}$
Bayesian Networks – Naive Bayes model

Unrealistic assumption:
Predictive variables dependent on effect variable.
Bayesian Networks – Noisy-MAX model

\[ P(Y_i | X_i) = c_i \]

State of \( D \) = logical MAX of \( Y_1..Y_N \)
Assignment

- Provide validated model for PD estimation for a credit portfolio. Logistic regression and Bayesian networks should be applied.
- Explore Bayesian network modeling and improve parameter learning techniques.
Remainder of Presentation

- Credit Rating
- Bayesian Networks
- Assignment
- Bayesian Network Learning Research:
  - Noisy-MAX learning algorithm
  - Prior / Posterior learning
- Bayesian Networks in Credit Rating
- Conclusions
Parameter learning based on observations not possible

$X_1 ... X_N$

$Y_1 ... Y_N$

$\gamma_1 ... \gamma_N$ unknown
Noisy-MAX Learning - Approach

- Construct noisy-MAX CPT for node $D$, based on nodes $X_1..X_N$ and parameters $c_1..c_N$
- Learn normal CPT (not noisy-MAX) for node $D$, based on $X_1..X_N$
- Minimize difference between the two CPT’s
Noisy-MAX Learning – Noisy-MAX CPT

CPT for node $Y_1$

<table>
<thead>
<tr>
<th></th>
<th>$X_1 = 2$</th>
<th>$X_1 = 1$</th>
<th>$X_1 = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_1 = 2$</td>
<td>$c_{122}$</td>
<td>$c_{121}$</td>
<td>$0$</td>
</tr>
<tr>
<td>$Y_1 = 1$</td>
<td>$c_{112}$</td>
<td>$c_{111}$</td>
<td>$0$</td>
</tr>
<tr>
<td>$Y_1 = 0$</td>
<td>$1 - c_{112} - c_{122}$</td>
<td>$1 - c_{111} - c_{121}$</td>
<td>$1$</td>
</tr>
</tbody>
</table>
Noisy-MAX Learning – Noisy-MAX CPT

Cumulative CPT for node $Y_1$

<table>
<thead>
<tr>
<th></th>
<th>$X_1 = 2$</th>
<th>$X_1 = 1$</th>
<th>$X_1 = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_1 = 2$</td>
<td>$c_{122}$</td>
<td>$c_{121}$</td>
<td>0</td>
</tr>
<tr>
<td>$Y_1 = 1$</td>
<td>$c_{112} + c_{122}$</td>
<td>$c_{111} + c_{121}$</td>
<td>0</td>
</tr>
<tr>
<td>$Y_1 = 0$</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

$p(Y_1 \geq y_1 \mid X_1 = x_1)$
Transformation of cumulative into normal:

\[ p(Y_1 = y_1 \mid X_1 = x_1) = p(Y_1 \geq y_1 \mid X_1 = x_1) - p(Y_1 \geq y_1 + 1 \mid X_1 = x_1) \]
For MAX gates, the following holds:

\[
p(D \geq d \mid X_1 = x_1, \ldots, X_N = x_N)
= p((Y_1 \geq d \mid X_1 = x_1) \lor (Y_2 \geq d \mid X_2 = x_2) \lor \ldots)
= 1 - \prod_{i=1}^{N} 1 - p(Y_i \geq d \mid X_i = x_i)
\]

→ Values for cumulative noisy-MAX CPT
Noisy-MAX Learning – Noisy-MAX CPT

Entries in noisy-MAX CPT defined as:

\[ \theta_{ij} = \prod_{x_p' \in X} \sum_{k=1}^{j} q_{prk} - \prod_{x_p' \in X} \sum_{k=1}^{j-1} q_{prk} \]

where

\[ \theta_{ij} = p(D = d_j | X = x_i) \]
\[ q_{ijk} = p(Y_i = y_k | X_i = x_j) \]
Noisy-MAX Learning – Normal CPT

Normal CPT obtained by parameter learning
Noisy-MAX Learning – Distance

- Euclidean Distance

\[ \sum_{i} \sum_{j} \left( \theta_{ij}^{CPT} - \theta_{ij}^{noisy-MAX} \right)^2 \]

- Kullback-Leibler Distance

\[ \sum_{i} \sum_{j} \theta_{ij}^{CPT} \ln \frac{\theta_{ij}^{CPT}}{\theta_{ij}^{noisy-MAX}} \]
Noisy-MAX Learning – Optimization

Gradient Descent:

\[\text{repeat} \]
\[\text{for each noisy-MAX parameter } c \text{ do} \]
\[c^* = c \pm \text{step} \]
\[d = \text{CalculateDistance} \]
\[\text{if } d < d_{\text{min}} \text{ then} \]
\[d_{\text{min}} = d \]
\[c = c^* \]
\[\text{end if} \]
\[\text{end for} \]
\[\text{until no distance reduction possible} \]
**Noisy-MAX Learning – Summary**

- Algorithm performs well
- Very fast
- Global minimum guaranteed
- EM performs better
- CPT not optimal starting point
CPT approach..

..but

- prior is initial value in CPT
- posterior is prior updated by observations
Prior / Posterior - Motive

CPT model has too many degrees of freedom
→ Rare entries have unreliable values
→ Resulting model is not stable
→ Setting priors can reduce instability
Prior / Posterior - Stages

1. Prior assessment
2. Posterior learning
3. Merging
CPT probabilities:

- **Beta distribution** (binary variables)
  \[
p(\theta | \alpha\beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}
\]

- **Dirichlet distribution** (multinomial variables)
  \[
p(\theta | \alpha_1..\alpha_N) = \frac{\Gamma\left(\sum_{i=1}^{N} \alpha_i\right)}{\prod_{i=1}^{N} \Gamma(\alpha_i)} \prod_{i=1}^{N} \theta_{\alpha_i-1}
\]
Expectation of Beta distribution:

\[ \int \theta \, \text{Beta}(\alpha \beta) \, d\theta = \frac{\alpha}{\alpha + \beta} \]

Expectation of Dirichlet distribution:

\[ \int \theta_i \, \text{Dir}(\alpha_1..\alpha_N) \, d\theta_i = \frac{\alpha_i}{\sum_j \alpha_j} \]
CPT entries for variable $X$ with states $x_1..x_N$:

$$\theta_{ij} = p(x_j | p_i) = \frac{o(p_i, x_j)}{\sum_{k=1}^{N} o(p_i, x_k)}$$

$p_i$ = i-th combination of parent states

$o(p_i, x_j)$ = number of observations $(p_i, x_j)$ in the data set.
Goal: Stable priors for every entry in the CPT

Solution:
1. Build naive Bayes model
2. Learn parameters of CPT’s
3. Perform inference to obtain $p(d | x_1...x_N)$
Prior / Posterior – Prior assessment

Inference in naive Bayes:

\[
p(d \mid x_1 \ldots x_N) = \frac{p(x_1 \ldots x_N \mid d) \cdot p(d)}{p(x_1 \ldots x_N)} = \frac{\prod_i p(x_i \mid d) \cdot p(d)}{\prod_i p(x_i)}
\]

\[
= \prod_i p(x_i \mid d) \cdot p(d)
\]

\[
= \prod_i \left( p(x_i \mid d) \cdot p(d) + p(x_i \mid \bar{d}) \cdot p(\bar{d}) \right)
\]
Merging of priors and posteriors: weighting

\[
\theta_{ij} = \frac{(w \cdot p_{ij}) + o(p_i, x_j)}{w + \sum_{k=1}^{j} o(p_i, x_k)}
\]

weight \( w \) = ‘number of observations’ for the prior \( p_{ij} \)
Prior / Posterior – Summary

+ Very stable for rare event data
+ Naive Bayes priors guarantee high performance
+ Posterior CPT learning provides flexibility
BN in Credit Rating – Project results

- Construction of data set
- Development of logistic regression model
- BN tests with C++ tool
- Development of BN models
- Algorithm for Noisy-MAX
- Design & implementation of prior / posterior learning
- Comparison of BN and logistic regression
## BN in Credit Rating - Comparison

<table>
<thead>
<tr>
<th>Model</th>
<th>GINI score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression</td>
<td>78.3%</td>
</tr>
<tr>
<td>BN – CPT</td>
<td>55.6%</td>
</tr>
<tr>
<td>BN – Noisy-MAX</td>
<td>75.4%</td>
</tr>
<tr>
<td>BN – Naive Bayes</td>
<td>77.5%</td>
</tr>
<tr>
<td>BN – Prior / Posterior</td>
<td>77.3%</td>
</tr>
</tbody>
</table>
BN in Credit Rating - Comparison

Logistic regression

- Favored by data selection process
- Optimization technique is by nature bound to yield good results
- Provides the right number of degrees of freedom
BN in Credit Rating - Comparison

Bayesian Networks

- Expert knowledge could not be incorporated
- CPT has too many degrees of freedom
- Naive Bayes performs well; apparently right number of degrees of freedom
- Logistic regression PD calculation and naive Bayes inference show similarities
Conclusions & Recommendations

BN is a very powerful modeling technique:

- Expert knowledge combined with statistical data
- High acceptance due to transparency
- Good performance in diagnosis

(media, trouble shooting, Microsoft Office Assistant)
Conclusions & Recommendations

Improvements for BN results:

→ Data selection process (stepwise BN)
→ High quality expert knowledge

Tax payment $C_1$

Financial setback $C_3$

Current Balance $X_1$

Extra cost $C_2$

Adverse market $C_4$

$PD$
Conclusions & Recommendations

CPT approximation for Noisy-MAX learning

- Noisy-MAX is very good alternative to CPT for rare event data
- CPT is probably not the best starting point for optimal parameters
Conclusions & Recommendations

Prior / Posterior learning

- Priors provide very stable networks
- High performance, promising technique
- More testing on data
Bayesian Networks in Credit Rating

Samuel Gerssen

March 12, 2004

Prof. Dr. H. Koppelaar
Drs. Dr. L.J.M. Rothkrantz
Ir. W. de Jong
Dr. K. van der Meer
Dr. D. Fokkema (ABN AMRO)
Dr. Ir. M.J. Druzdzel (University of Pittsburgh)