Building a Financial Case-Based Reasoning Prototype from Scratch with Respect to Credit Lending and Association Models driven by Knowledge Discovery

Jürgen Hönigl ¹  Yuliya Nebylovych ²

¹Institute for Application-Oriented Knowledge Processing
Johannes Kepler University
Linz, Austria

²Department for Applied Linguistics
Lviv Polytechnic National University
Ukraine

November 1, 2012
Case

Def. I
Problem, solution and annotations

Def. II
(i) a situation and its goal, (ii) the solution and, sometimes, means of deriving it, (iii) the result of carrying it out, (iv) explanations of results, and (v) lessons that can be learned from the experience.

Def. III
contextualized piece of knowledge representing an experience that teaches a lesson fundamental to achieving the goals of the reasoner

\[1\]

\( R^4 \)

- Aamodt / Plaza
- Retrieve, Reuse, Revise and Retain
### Alpha/Beta Error

<table>
<thead>
<tr>
<th>Solvency Assumed</th>
<th>Solvency Actually</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>Bad</td>
<td>Beta Error</td>
</tr>
<tr>
<td>Good</td>
<td>Accuracy</td>
</tr>
<tr>
<td>Bad</td>
<td>Alpha Error</td>
</tr>
</tbody>
</table>
German Credit Data Set - Pre-processing (excerpt)

Attribute 10 describes qualitative the other debtors and guarantors.

\[
att_{10} = \begin{cases} 
A101 : & \text{none} \\
A102 : & \text{co-applicant} \\
A103 : & \text{guarantor}
\end{cases}
\]
Problem / Solution

Problem - given query
Problem = Age, Credit Amount, Credit History, Duration, Income, Purpose

Problem with 2 additional attributes
Problem = Age, Credit Amount, Credit History, Duration, Income, Married, Other Debtors Guarantors, Purpose

Solution
Solution = Cost Factor, Recommendation
Hotspot

Waikato Environment for Knowledge Analysis

About
HotSpot learns a set of rules (displayed in a tree-like structure) that maximize/minimize a target variable/value of interest.

[^3]: http://weka.sourceforge.net/packageMetaData/hotSpot/index.html accessed on 20120827
Example (I)

Cost_Factor=A211 (70% [700/1000])
  Status_Checking_Account = A14 (88.32% [348/394])
  |  Credit_Amount <= 7824 (89.67% [330/368])
  Duration_Months <= 15 (79.35% [342/431])
  |  Credit_Amount <= 3973 (80.88% [330/408])
  Age > 34 (76.11% [344/452])
  |  Credit_Amount <= 9436 (77.44% [333/430])

4  

Branching factor:3, Segment size: 33%
Example (II)

```
Cost_Factor=A212 (30% [300/1000])
Duration_Months > 8 (32.01% [290/906])
Foreign_Worker = A201 (30.74% [296/963])
  | Credit_Amount > 601 (31.17% [293/940])
Age <= 61 (30.46% [293/962])
  | Credit_Amount > 601 (30.88% [290/939])
```

Branching factor:3, Segment size: 29%
Example (III)

\[
\begin{align*}
\text{Cost\_Factor} &= A211 \ (70\% \ [700/1000]) \\
\text{Credit\_Amount} &\leq 3913 \ (74.46\% \ [551/740]) \\
\text{Duration\_Months} &\leq 26 \ (74.32\% \ [573/771]) \\
&\mid \quad \text{Credit\_Amount} \leq 6419 \ (75.37\% \ [554/735]) \\
\text{Age} &> 26 \ (72.89\% \ [554/760]) \\
&\mid \quad \text{Credit\_Amount} \leq 12204 \ (73.83\% \ [550/745]) \\
\text{Other\_Installment\_Plans} &= A143 \ (72.48\% \ [590/814]) \\
&\mid \quad \text{Duration\_Months} < 42 \ (74.54\% \ [568/762]) \\
&\quad \mid \quad \text{Credit\_Amount} \leq 7814 \ (76.59\% \ [553/722]) \\
&\quad \mid \quad \text{Credit\_Amount} \leq 7166 \ (74.53\% \ [550/738]) \\
\text{Age} &> 22 \ (73.3\% \ [560/764]) \\
&\quad \mid \quad \text{Credit\_Amount} \leq 10875 \ (74.66\% \ [554/742])
\end{align*}
\]

\(^6\)Branching factor: 4, Segment size: 55%
Example (IV)

Credit_History=A32 (53% [530/1000])
   Number_of_Existing_Credits_at_this_Bank <= 1 (75.51% [478/633])
      Duration_Months <= 36 (77.66% [445/573])
      Credit_Amount <= 7408 (76.91% [443/576])
      Age <= 52 (76.43% [441/577])
   Credit_Amount <= 4675 (55.9% [445/796])
      Number_of_Existing_Credits_at_this_Bank <= 2 (57.35% [441/769])
      Other_Installment_Plans = A143 (55.53% [452/814])
      Number_of_Existing_Credits_at_this_Bank <= 2 (56.96% [450/790])

7 Branching factor: 3, Segment size: 33%
Keywords - Side Effect

- Amount
- Age
- Income
- Duration
Conclusion

- **Good**
  - $R^4$ model
  - suitable arguments for associations

- **Bad**
  - too less knowledge discovery
    - application
    - approach

- **Ugly**
  - inappropriate static rules
  - many similar redundant cases within the case base

---

*8 The Good, the Bad and the Ugly (1966), Il buono, il brutto, il cattivo. (original title) by Sergio Leone, 1966 (USA), United Artists*
Future Work

- similarity measures
- different attributes
- try and test weights
at the end

Any questions?
Thanks for listening.