

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

Jürgen Hönigl<sup>1</sup>    Yuliya Nebylovykh<sup>2</sup>

<sup>1</sup>Institute for Application-Oriented Knowledge Processing  
Johannes Kepler University  
Linz, Austria

<sup>2</sup>Department for Applied Linguistics  
Lviv Polytechnic National University  
Ukraine



Software Engineering  
Conference in Russia

# 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

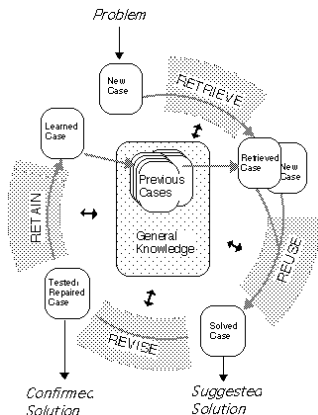
---

<sup>1</sup>II & III: Ralph Bergmann, Janet L. Kolodner, and Eric Plaza.

Representation in case-based reasoning. Knowledge Eng. Review, 20(3):209213, 2005.

R<sup>4</sup>

- Aamodt / Plaza
- Retrieve, Reuse, Revise and Retain



2

<sup>2</sup>Agnar Aamodt and Eric Plaza. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Commun.*, 7(1):3959, 1994.

## Alpha/Beta Error

	Solvency Actually	
Solvency Assumed	Good	Bad
Good	Accuracy	Alpha Error
Bad	Beta Error	Accuracy

## 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

---

<sup>3</sup><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

---

<sup>4</sup>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])
```

5

---

<sup>5</sup>Branching factor:3, Segment size: 29%

## Example (III)

```
Cost_Factor=A211 (70% [700/1000])
  Credit_Amount <= 3913 (74.46% [551/740])
  Duration_Months <= 26 (74.32% [573/771])
    | Credit_Amount <= 6419 (75.37% [554/735])
  Age > 26 (72.89% [554/760])
    | Credit_Amount <= 12204 (73.83% [550/745])
  Other_Installment_Plans = A143 (72.48% [590/814])
    | Duration_Months <= 42 (74.54% [568/762])
      | Credit_Amount <= 7814 (76.59% [553/722])
    | Credit_Amount <= 7166 (74.53% [550/738])
    | Age > 22 (73.3% [560/764])
      | Credit_Amount <= 10875 (74.66% [554/742])
```

6

---

<sup>6</sup>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

---

<sup>7</sup>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

---

<sup>8</sup>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?



eof

Thanks for listening.