A Smart Financial Advisory System exploiting Case-Based Reasoning

Giorgio Leonardi, Luigi Portinale, Paolo Artusio and Marco Valsania

Istituto di Informatica, Università del Piemonte Orientale, Alessandria, Italy
ORS Srl, Roddi (CN), Italy
The project SmartFasi

• Funded by the ICT Innovation cluster of Piedimont Region

• Web-based financial decision support system, composed (mainly) by:
  – Cloud and HPC infrastructure for High-intensity activities
  – Stochastic modeling, Montecarlo simulations and CBR
    • Simulation of asset behaviour in a specific time horizon
The project SmartFasi: CBR

- Simulation tools leave the user alone in the choice of financial products

- CBR recommendation system:
  - Knowledge-based recommendation strategy
  - Underlying assumption: individuals who share similar financial needs will act on the market in a similar way
  - Extension of the asset picking phase showing to the user the products selected by customers having similar personal and financial profile
The project SmartFasi: CBR

• Addresses the following targets for different end-users

• Private Investors
  – Improving the vision of the global scenario showing how similar users act on the market

• Professional Users:
  – Proposing to the customers investment scenarios tailored on the customer’s profile, thus personalizing the service
  – Benchmarks can be used to compare scenarios more suitable to the customer
  – Performing historical analyses on cluster of clients to discover trends of investments to be supported or contrasted for the customer
  – Improving customer acquisition process by tying business targets to the interests/characteristics of the customers
The SmartFasi recommendation module

• Case definition:
  – Customer’s personal and financial characteristics
  – Solution: recommendation of financial products more popular among similar customers

• Proper distance measures between customers

• Multi-step recommendation strategy:
  – Query Q: Target Customer case (TC)

---

Step 1
personal data POV

Selection of the most similar customers wrt to the target one

Step 2
Financial POV

Further filter on the most similar portfolios of the selected customers

Step 3
Final result

Extraction of the most successful products in the selected portfolios
Case definition

- Case defined as features of a customer:
  - Personal data
  - Investment capabilities and financial adequacy
  - Composition of portfolio(s)

<table>
<thead>
<tr>
<th>Feature #</th>
<th>Feature</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Age</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>Available capital</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>Risk tolerance / Adequacy</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>Asset allocation for each portfolio</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>N. of children</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>Marital status</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>Education</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>Sex</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>Type of employment</td>
<td>1</td>
</tr>
</tbody>
</table>
Case definition: 2-level portfolios

- **Level 1**: Percentage of assets in each of the 6 categories
  - Defines the general investment strategy → used in Step 1 together with personal data
- **Level 2**: Percentage of each specific asset (S0..S59)
  - Describes in detail the financial behaviour → used in Step 2

![Diagram showing 2-level portfolios with categories and specific assets](image-url)
Step 1

- Selection of the $N$ most similar customers wrt the target one
- Nearest Neighbor Search using information from Table 1 (personal and portfolios at level 1)
- HEOM distance measure adopted

$$D_{HEOM}^f (x, y) = \begin{cases} 
1 & \text{if } x \text{ or } y \text{ is unknown} \\
\text{overlap}(x, y) & \text{if } f \text{ is nominal} \\
\text{rn_diff}(x, y) & \text{otherwise}
\end{cases}$$

\[
\text{overlap}(x, y) = 0 \text{ if } x = y \text{ and } \text{overlap}(x, y) = 1 \text{ if } x \neq y
\]

\[
\text{rn_diff}(x, y) = \frac{|x-y|}{\text{range}(f)}
\]
Step 1: Distance between portfolios

- Portfolios are arrays and are treated as a mini-case

- Natural choice for distance between two arrays: cosine distance
  - Successfully adopted in the past in financial recommender systems

\[ D_{\text{cos}}(a, b) = 1 - \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2} \sqrt{\sum_{i=1}^{n} b_i^2}} \]

- Customers we are comparing might have a different number of portfolios
  - Need to define the notion of «best match» between those sets of portfolios
Step 1: Distance between portfolios

- Let $P_t$ and $P_c$ be the set of portfolios of the target and of a given customer

\[ P_t = (p^1_t, p^2_t, \ldots, p^n_t) \text{ and } P_c = (p^1_c, p^2_c, \ldots, p^m_c) \]

- Best match: pair of permutations of $P_t$ and $P_c$ giving the minimum overall distance between the portfolios in them:

\[ D_p(P_c, P_t) = \min_{\text{Perm}(P_c) \times \text{Perm}(P_t)} \frac{\sum_{i=1}^{\min(n,m)} D_{\cos}(p^i_c, p^i_t)}{\min(n,m)} \]

- The best matching portfolios are furtherly stored to be re-used in Step 2
Step 1: Overall distance between customers

- The overall distance is computed in order to list the N customers more similar to the target.

\[
D(C_1, C_2) = \frac{\sum_{i=1}^{s} w_i \cdot D(v_1^i, v_2^i)}{\sum_{i=1}^{s} w_i}
\]

where:

\[
D(v_1^i, v_2^i) = \begin{cases} 
D_p(v_1^i, v_2^i) & \text{if } i = 4 \text{ (portfolios)} \\
D_{HEOM}^i(v_1^i, v_2^i) & \text{otherwise}
\end{cases}
\]

- Similarity defined as: \( S(C_1, C_2) = 1 - D(C_1, C_2) \)
Step 2

• Receives the N customers from Step 1, together with the relative best matching portfolios

• Focus on the financial POV using portfolios at level 2
  – Cosine distance on the best matching portfolios and filtering of the J best matches more similar to the portfolios of the customer
  – The J selected portfolios are passed to Step 3

• Optional step: if the target customer has no portfolios yet, Step 2 is not applicable
  – All the portfolios of the N customers selected at Step 1 are passed to Step 3
Step 3

• Step 3 relies on a set of selected portfolios which:
  - Belong to customers with the same personal and overall financial characteristics (Step 1)
  - Are the closest wrt the target customer’s investment strategies, if she/he has one or more portfolios (Step 2)

• Step 3 scans the selected portfolios and lists the K most popular (or less popular) assets in them

• Retrieved assets are firstly filtered to remove those which are not compliant with the TC’s MiFID profile
Step 3: Statistics

• Each retrieved asset is associated with statistics, helping users providing more informed suggestions
  – Frequency
  – Avg. percentage
  – Avg. Distance of customers
  – Avg. Distance of portfolios

• Frequency is particularly important
  – Private investors: probably can trust more the assets popular among similar customers
  – Professional users: analyzing the set of popular products, the offer can be differentiated making the resulting portfolio more desirable
Case Study 1

• **Query: Target Customer 1:**
  Sex = M; Age = 41; No. of children = 2; Marital status = Married; Education = Undefined; Type of employment = freelancer; Adequacy = medium; Available Capital = 10,000 euros.

• **Owner of portfolio P1 (at Level 1):**
  BCF = 11%; BCV = 0%; BGF = 0.1%; BGV = 88.89%; BCC = 0%; BGC = 0%

• **Investment preferences:**
  – Many floating rate assets, but issued by government organizations
  – A small part of corporate assets, but with a fixed rate
  – Even if going towards a certain level of risk, some factors mitigates the risk itself.
Results: retrieved assets

<table>
<thead>
<tr>
<th>Title ID</th>
<th>F</th>
<th>AP</th>
<th>ADC</th>
<th>ADP</th>
<th>Asset class</th>
<th>Asset risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>411</td>
<td>0.6</td>
<td>41.0%</td>
<td>0.40242228</td>
<td>0.7855223</td>
<td>BCF</td>
<td>medium</td>
</tr>
<tr>
<td>394</td>
<td>0.6</td>
<td>3.0%</td>
<td>0.41156575</td>
<td>0.1189378</td>
<td>BGF</td>
<td>low</td>
</tr>
<tr>
<td>461</td>
<td>0.6</td>
<td>1.2%</td>
<td>0.38838637</td>
<td>0.6667488</td>
<td>BGF</td>
<td>low</td>
</tr>
<tr>
<td>583</td>
<td>0.4</td>
<td>33.3%</td>
<td>0.39654094</td>
<td>0.0001233</td>
<td>BGV</td>
<td>low</td>
</tr>
<tr>
<td>650</td>
<td>0.4</td>
<td>3.7%</td>
<td>0.39654094</td>
<td>0.0001233</td>
<td>BGV</td>
<td>low</td>
</tr>
<tr>
<td>5994</td>
<td>0.4</td>
<td>10.0%</td>
<td>0.38282567</td>
<td>0.7019767</td>
<td>BGV</td>
<td>low</td>
</tr>
<tr>
<td>9380</td>
<td>0.4</td>
<td>9.9%</td>
<td>0.38282567</td>
<td>0.7019767</td>
<td>BGV</td>
<td>low</td>
</tr>
<tr>
<td>9640</td>
<td>0.4</td>
<td>9.7%</td>
<td>0.38282567</td>
<td>0.7019767</td>
<td>BGV</td>
<td>low</td>
</tr>
<tr>
<td>9741</td>
<td>0.4</td>
<td>9.6%</td>
<td>0.38282567</td>
<td>0.7019767</td>
<td>BGV</td>
<td>low</td>
</tr>
<tr>
<td>9794</td>
<td>0.4</td>
<td>9.9%</td>
<td>0.38282567</td>
<td>0.7019767</td>
<td>BGV</td>
<td>low</td>
</tr>
<tr>
<td>68726</td>
<td>0.4</td>
<td>9.5%</td>
<td>0.38282567</td>
<td>0.7019767</td>
<td>BGV</td>
<td>low</td>
</tr>
<tr>
<td>71586</td>
<td>0.4</td>
<td>3.9%</td>
<td>0.39654094</td>
<td>0.0001233</td>
<td>BGF</td>
<td>low</td>
</tr>
</tbody>
</table>

- Recommended asset classes follow very closely the type of investments made by the TC.
  - 8 of them are titles issued by government with floating rates
  - 1 is a corporate title with fixed rate
- The system also recommends 3 government bonds at fixed rate.
  - Although they do not belong exactly to the same classes owned by the TC, they can represent a suitable suggestion.
Case Study 2

• **Query: Target Customer 2:**
  Sex = M; Age = 42; No. of children = 0; Marital status = Undefined; Education = Undefined; Type of employment = Education; Adequacy = medium; Available Capital = 12,000 euros.

• **Owner of portfolio P1 (at Level 1):**
  BCF= 0%; BCV= 0%; BGF= 51%; BGV= 49%; BCC= 0%; BGC= 0%

• **Investment preferences:**
  – Completely composed by government issued titles,
  – Distributed equally between fixed and variable rate
Results: retrieved assets

<table>
<thead>
<tr>
<th>Title ID</th>
<th>F</th>
<th>AP</th>
<th>ADC</th>
<th>ADP</th>
<th>Asset class</th>
<th>Asset risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>1645</td>
<td>0.6</td>
<td>1.4%</td>
<td>0.37407598</td>
<td>0.1349692</td>
<td>BGF</td>
<td>low</td>
</tr>
<tr>
<td>411</td>
<td>0.4</td>
<td>11.2%</td>
<td>0.3900423</td>
<td>0.0002465</td>
<td>BCF</td>
<td>medium</td>
</tr>
<tr>
<td>461</td>
<td>0.4</td>
<td>0.1%</td>
<td>0.3900423</td>
<td>0.0002465</td>
<td>BGF</td>
<td>low</td>
</tr>
<tr>
<td>1228</td>
<td>0.4</td>
<td>6.9%</td>
<td>0.36215407</td>
<td>0.0000005</td>
<td>BGV</td>
<td>low</td>
</tr>
<tr>
<td>1693</td>
<td>0.4</td>
<td>99.9%</td>
<td>0.36215407</td>
<td>0.0000005</td>
<td>BGF</td>
<td>low</td>
</tr>
<tr>
<td>5994</td>
<td>0.4</td>
<td>10.0%</td>
<td>0.3900423</td>
<td>0.0002465</td>
<td>BGV</td>
<td>low</td>
</tr>
<tr>
<td>9380</td>
<td>0.4</td>
<td>9.9%</td>
<td>0.3900423</td>
<td>0.0002465</td>
<td>BGV</td>
<td>low</td>
</tr>
<tr>
<td>9640</td>
<td>0.4</td>
<td>9.7%</td>
<td>0.3900423</td>
<td>0.0002465</td>
<td>BGV</td>
<td>low</td>
</tr>
<tr>
<td>9741</td>
<td>0.4</td>
<td>9.6%</td>
<td>0.3900423</td>
<td>0.0002465</td>
<td>BGV</td>
<td>low</td>
</tr>
<tr>
<td>9794</td>
<td>0.4</td>
<td>9.9%</td>
<td>0.3900423</td>
<td>0.0002465</td>
<td>BGV</td>
<td>low</td>
</tr>
<tr>
<td>68726</td>
<td>0.4</td>
<td>9.5%</td>
<td>0.3900423</td>
<td>0.0002465</td>
<td>BGV</td>
<td>low</td>
</tr>
<tr>
<td>76979</td>
<td>0.4</td>
<td>9.5%</td>
<td>0.3900423</td>
<td>0.0002465</td>
<td>BGV</td>
<td>low</td>
</tr>
</tbody>
</table>

- Recommended assets are mostly government issued with fixed and variable rate.
- Except for asset 411: a corporate title with fixed rate and medium risk level. It differs from the investments of the TC, but:
  - It is a more risky asset, but having features (the fixed rate) somewhat mitigating the risk.
- Asset 411 could diversify the investments of the TC without sacrificing her/his overall investment strategies.
Discussion: improving the «standard» asset baskets

- Identify a representative customer for each cluster and run our engine
- The most (or less) popular assets retrieved are used to create alternative baskets (BIUs) tailored on the actual behaviour of the customers
- BIUs may be used to update the original baskets, eventually upgrading their value making them more appealing for new customers
- Historical analysis of BIUs helps discovering investment trends inside each cluster
Conclusion and future work

- SmartFasi contains our CBR recommendation strategy, defining:
  - The notion of a customer and of her/his investment preferences
  - A suitable notion of similarity among customers
  - A multi-step procedure to retrieve the most (or less) popular items among similar customers
  - Module complementary to the analytical engine of SmartFasi (Montecarlo simulations)

- Recommended titles can be usefully exploited by:
  - Private investors
  - Professional users

- CBR also useful for making the asset baskets more appealing for new customers

- In the future: set up experimental plans to deeply evaluate the strategies exposed
Thank you very much!