Personalized Finance Advisory through Case-based Recommender Systems and Diversification strategies*

Speaker:
PhD. Student Marco Polignano

*Cataldo Musto, Marco Polignano, Giovanni Semeraro, Pasquale Lops, Marco de Gemmis, Università degli Studi di Bari "Aldo Moro", Department of Computer Science; Georgios Lekkas Objectway Financial Software, Italy
• **Common situation**
  – One model portfolio for each class of investor hand made by Advisory Desk Office, small modifications of the model by Financial Advisors from the analysis of MiFid questionnaire responses

• **Target situation**
  – Interactive system to recommend a list of personalized financial portfolios in according to the user targets and past useful portfolios collected from the broker’s investments
Research purposes

• Introduces a novel framework for financial products recommendations

• Evaluates the effectiveness of CBRS recommendation strategies in a special (and, up to our knowledge, not yet evaluated) domain

• Proposes a greedy diversification algorithm able to diversifying the investment strategies over time

• Evaluates the effectiveness of the framework through an extensive ex-post evaluation

"Investing is putting out money to be sure of getting more money back later at an appropriate rate."  
Warren Buffett
Case-based Recommender Systems*

• Subcategory of *knowledge-based recommender systems*

• Case-based Recommender Systems workflow:
  – Explicate the user needs
  – Access to the case base repository
  – Set the system restrictions and get useful historical cases
  – Show the recommendations

• Pros
  – *No hand-coded rules*

• Cons
  – *Needs a big dataset of historical cases*

Case-based Recommendation Pipeline

- **Case-Base Update (Retain)**
- **Neighborhoods Identification (Retrieve)**
- **Extraction of Candidate Portfolios (Reuse)**
- **Further Discussion (Review)**

**Case base**

- **User Properties**
- **Final Portfolio**
- **Candidate Portfolios**
- **Ranked Portfolios**

**Further Discussion**

**Case Base Update**

**Neighborhoods**

**Final Portfolio**

**Candidate Portfolios**

**Ranked Portfolios**

**Extraction of Candidate Portfolios**

**Ranking of Candidate Portfolios**

**Case-based Recommendation Pipeline**

- **Personalized Finance Advisory through Case-based Recommender Systems and Diversification strategies**
Case base in a financial context

\[ C \in CB \subseteq I \times U \times E \]

\[ C = (i, u, e) \]

- **i \in I**: Each portfolio is represented as the distribution of the asset classes that compose it, such as Euro Bond, High Yield Bond, Emerging Markets Stock Options and so on, along with their percentage.

<table>
<thead>
<tr>
<th>Euro Bond</th>
<th>North America Stocks</th>
<th>Pacific Stocks</th>
<th>Global Stocks</th>
<th>High risk stocks</th>
<th>Low risk stocks</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>20%</td>
<td>24%</td>
<td>15%</td>
<td>13%</td>
<td>12%</td>
</tr>
</tbody>
</table>

- **u \in U**: Each user \( u_i \) is represented as a vector of eight features:
  - risk profile, investment goals, temporal goals, financial experience, financial situation, advice type, sex and age.

- **e \in E**: recommendations evaluating model \( \cong \text{Yield of each portfolio} \) in the past investments
Retrieve and Reuse

Identify neighbourhoods with similar (already solved) cases which can be potentially useful to tailor the investment proposal.

• This first part can be tackled as a classical retrieval task.
  – Given a vector-space representation of the target user the process returns a set of k similar users \( N = n_1 \cdots n_k \)

• Two different approaches have been implemented:
  – **User match**: retrieves all the cases whose users exactly share the same features
  – **Cosine similarity**: Let \( \vec{u}_i \) be a vector space representation of the target user and \( \vec{c}_i \) be a vector space representation of another user already stored in the case base, cosine similarity can be computed as follows:

\[
\cos(\vec{u}, \vec{c}) = \frac{\sum_{i=1}^{n} \vec{u}_i \vec{c}_i}{\sqrt{\sum_{i=1}^{n} (\vec{u}_i)^2} \sqrt{\sum_{i=1}^{n} (\vec{c}_i)^2}}
\]
Revise, Review and Retain

The proposed list of recommended portfolios is revised by adopting one of the five different revise techniques implemented:

1. **Basic Ranking**: ranked according to the output produced by the Retrieve step

2. **Clustering**: retrieved portfolios are clustered according to the $k$-means clustering algorithm

3. **Greedy Diversification**: the approach tries to diversify the final solutions by iteratively picking from the original set of candidate solutions the ones with the best trade-off between similarity and diversity

4. **FCV**: Financial Confidence Value (FCV) calculates how close to the optimal one is the distribution of the asset classes in a portfolio, according to the average historical yield obtained by each class

5. **FCV + Greedy**: this combined strategy first uses the greedy algorithm to diversify the solutions, then exploits FCV to rank the portfolios and obtain the final solutions

Human advisor and investor can further discuss and modify the portfolio selected from the list of recommended in order to come to the final solution, this solution will be stored in the case base.
Revise: Greedy Diversification*

Let $u$ be the target user, $F$ be the set of final solutions, $C_{retr}$ be a set of previously retrieved cases, at each step the algorithm ranks the retrieved neighbours by calculating the quality score of each case $c_i = (u_i, p_i, f_i) \in C_{retr}$ as follows:

$$Quality(u, c_i, F) = \cos(u, u_i) \times relDiv(p_i, F)$$

$$relDiv(p_i, F) = \sum_{j=1}^{\left|F\right|} \frac{1 - \cos(s_i, f_j)}{\left|C\right|}$$

At each step, the solution with the highest quality score is removed from the set of candidate solutions and is stored in $F$.

At the first iteration $F$ is empty, $relDiv(n; F) = 1$ for all the neighbours. Thus, the first item selected is the one with the highest similarity. Next, at each iteration, the solution with the best score is chosen.

Revise: FCV Ranking

We adopt the **Interest Confidence Value (ICV)**\(^*\) to the financial domain.

Given a set of asset classes \(A\), for each portfolio \(p\), the set \(P\) of the asset classes **which compose it** and its complement \(\bar{P}\), are computed.

**FCV** is formally defined as:

\[
FCV(p) = Y(p)^{\log(\lambda) + 1}
\]

\[
Y(p) = \sum_{i=1}^{\vert P \vert} p_{a_i} * y_{a_i}
\]

\[
\lambda = \frac{\sum_{i=1}^{\vert P \vert} y_{a_i}}{\sum_{k=1}^{\vert \bar{P} \vert} y_{a_k}}
\]

\(Y(p)\) is the **total yield obtained by the portfolio**, and \(\lambda\) is a **drift factor** which calculates the ratio in terms of average yield between the asset classes in the portfolio and those which are not in.

For values of \(\lambda \geq 1\), it acts as a **boosting** factor (for \(\lambda << 1\), it acts as **dumping** factor)

Experiment goal

• **GOAL 1**
  Analysis of the **influence of each parameter** of the framework (similarity measure, feature combinations, revise and diversification strategies) **on the performance** of recommended portfolios

• **GOAL 2**
  **Comparison** of the **performance of recommended portfolios** to that of **portfolios proposed by a human advisor**

• **GOAL 3**
  **Ex-post comparison of the best-performing configuration** to the portfolios proposed by a human advisor after three and after six months from the agreement date.
Experiment Configuration

- **Dataset**: 1172 real anonymous users who agreed portfolios with financial advisors between June 2011 and June 2013

- **Portfolio composition**: 19 different asset classes along with their percentage

- **Yield time-slice**: yield generated by each portfolio from the agreement date to January 2014.

- **Similarity Measures**: *User Match and Cosine Similarity*

- **Set of features**: *basic* (features 1-5, including all the financial-based features), *extended* (features 1-6, it adds the advice type to the basic set) and *complete* (features 1-8, it adds the demographic information to the extended set).

<table>
<thead>
<tr>
<th>#id</th>
<th>feature</th>
<th>type</th>
<th>domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>risk profile</td>
<td>ordinal</td>
<td>[very low, low, normal, high, very high]</td>
</tr>
<tr>
<td>2</td>
<td>investment goals</td>
<td>ordinal</td>
<td>[very low, low, normal, high, very high]</td>
</tr>
<tr>
<td>3</td>
<td>temporal goals</td>
<td>ordinal</td>
<td>[very early, early, normal, long, very long]</td>
</tr>
<tr>
<td>4</td>
<td>financial experience</td>
<td>ordinal</td>
<td>[very low, low, normal, high, very high]</td>
</tr>
<tr>
<td>5</td>
<td>financial situation</td>
<td>ordinal</td>
<td>[very low, low, normal, high, very high]</td>
</tr>
<tr>
<td>6</td>
<td>advice type</td>
<td>nominal</td>
<td>[normal, extended]</td>
</tr>
<tr>
<td>7</td>
<td>sex</td>
<td>nominal</td>
<td>{male, female}</td>
</tr>
<tr>
<td>8</td>
<td>age</td>
<td>integer</td>
<td>[18...80]</td>
</tr>
</tbody>
</table>
Experiment 1: Influence of Parameters

Comparing the best performing configuration obtained by user match and cosine similarity strategies (CS-Complete vs UM-Basic), geometrical retrieval strategy significantly outperform user matching in terms of average yield.
Experiment 1: Influence of Parameters

Comparing the best performing configuration obtained by user match and cosine similarity strategies (CS-Complete vs UM-Basic), geometrical retrieval strategy significantly outperform user matching in terms of average yield.
Experiment 1: Influence of Parameters

By analysing the results, it emerges that the **best-performing configuration** is the one based on **cosine similarity and FCV ranking strategy**, with a richer case representation based on all the available features.
Experiment 1: Influence of Parameters

Results:

- The adoption of Cosine Similarity retrieval results in a statistically significant improvement of approximately 10% with respect to User Match retrieval.

- The configuration which provides the best diversity is the Greedy.

- The configuration which provides the best average yield is the FCV.

- The FCV + Greedy strategy is able to lead to diversified recommendations which can provide the user with good average yield as well.
Experiment 2: Comparison to Baselines

All the approaches adopted, **significantly outperform** Human and Collaborative Recommendation baselines about average yield of recommended portfolios.
Experiment 3: Ex-post evaluation (January-April 2014)
Experiment 3: Ex-post evaluation (January-July 2014)
Conclusions and Future Work

• Our approach integrates a novel strategy based on a Greedy algorithm aiming at diversifying investment proposals.

• Experiments performed on a dataset of 1172 real users provided several interesting outcomes, since it emerged that the proposed approach can significantly outperform both a baseline represented by an Item-based CF and the recommendations provided by a human advisor.

• The ex-post evaluation at three and six months further confirmed these results, since our strategy leads to both diversified and fruitful investment proposals.

• Future work:
  – Evolve our recommendation approach in a conversational model making the advisor able to concretely discuss with the recommender systems
  – Extend the representation of the cases introducing novel features
Any Questions?