# Financial Product Recommendation through Case-based Reasoning and Diversification Techniques

Cataldo Musto, Giovanni Semeraro,
Pasquale Lops, Marco De Gemmis
University of Bari Aldo Moro, Italy
{name.surname}@uniba.it

Georgios Lekkas
Objectway Financial Software
http://www.objectway.com/
lekkas@objectway.com

## ABSTRACT

This work[1] proposes a framework for financial product recommendation which combines *case-based reasoning* with diversification techniques to support wealth managers in recommending personalized investment portfolios. The performance of the framework has been evaluated against 1172 real users, and results show that the yield obtained by recommended portfolios overcomes that of portfolios proposed by human advisors in many experimental settings.

## 1. MOTIVATIONS AND METHODOLOGY

Widespread recommendation approaches, such as content-based (CB) and collaborative filtering (CF), can hardly put into practice in the domain of *financial recommendations*. Typically, pure CB strategies are likely to fail since content information describing both users and financial products is too poor to feed a CB recommendation algorithm. Moreover, the *over-specialization problem*, typical of CB recommenders, collide with the fact that turbulence and fluctuations in financial markets suggest to change and diversify the investments over time. On the other side, CF algorithms may lead to the problem of *flocking*, since *user-based CF* could move many (similar) users to invest in the same asset classes at the same time, making the algorithm victim of potential trader attacks[2].

As a consequence, we focused the attention on different recommendation paradigms. Given that *financial advisors* have to analyze and sift through several *investment portfolios*[3] before providing the user with a solution able to meet his investment goals, the insight behind our recommendation framework is to exploit case-based reasoning (CBR) to tailor investment proposals on the ground of a *case base* of previously proposed investments. Formally, given a case library $C$, each case $c_i \in C$ is a triple $\langle u_i, p_i, f_i \rangle$, where $u_i$ is a representation of a user, $p_i$ is a representation of the portfolio, and $f_i$ is a feedback assessment. Each user $u_i$ is represented as a vector of five features: *risk profile*[4], inferred through the standard MiFiD[5] questionnaire, *investment goals*, *temporal goals*, *financial experience* and *financial situation*. Our recommendation process is based on the typical CBR workflow, described in [1], and is stuctured in three different steps:

**(1) Retrieve and Reuse:** retrieval of similar portfolios is performed by representing each user as a vector according to the weight of each feature (very low=1, very high=5). Next, *cosine similarity* is adopted to retrieve the most similar users (along with the portfolios they agreed) from the case base.

**(2) Revise:** the *candidate solutions* retrieved by the first step are typically too many to be consulted by a human advisor. Thus, the REVISE step further filters this set to obtain the *final solutions*. Five techniques have been introduced for the revise of the list:

**(a) Basic Ranking:** portfolios are ranked in descending *cosine similarity* order, according to the scores returned by the RETRIEVE step. The first $k$ portfolios are returned to the advisor as *final solutions*.

**(b) Greedy Diversification:** this strategy implements the diversification algorithm described in [3]. The algorithm tries to diversify the *final solutions* by iteratively picking from the original set of *candidate solutions* the ones with the best compromise between *cosine similarity* and *intra-list diversity* with respect to the previously picked solutions. At each step of the strategy, the solution with the best compromise is removed from the set of candidate solutions and is stored in the set of final solutions.

**(c) FCV**: this strategy adapts the *Interest Confidence Value* proposed in [2] to the financial domain. *Financial Confidence Value (FCV)* calculates how close to the optimal one is the distribution of the asset classes in a portofolio, according to the average historical yield obtained by each class. Given a set of asset classes $A$, for each portfolio $p$ the set $P$, which contains the asset classes which compose it, and its complement $\overline{P}$ are computed. Next, FCV is formally defined as:

$$FCV(p) = Y(p)^{log(\lambda)+1} \tag{1}$$

[1]This work fullfils the research objectives of the projects ObjectWay-Finance-as-a-Service: Smart Application software and Service for Financial Services Operators and PON 01 00850 ASK-Health (Advanced System for the interpretation and sharing of knowledge in health care).

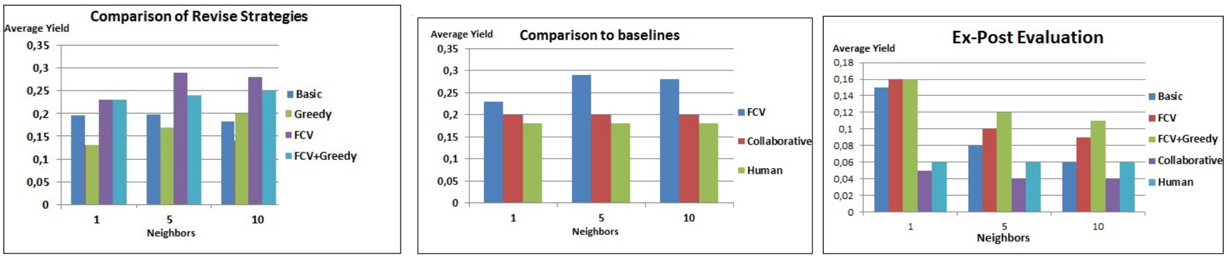
[2]http://www.technologyreview.com/view/425654/flocking-behaviour-improves-performance-of-financial-traders/

[3]http://en.m.wikipedia.org/wiki/Portfolio_(finance)

[4]http://www.investopedia.com/terms/r/risk-profile.asp

[5]http://en.wikipedia.org/wiki/Markets_in_Financial_Instruments_Directive


RecSys 2014 Poster Proceedings, October 6-10, 2014, Foster City, Silicon Valley, USA.




(a) Comparison of revise strategies (b) Comparison to baselines (c) Ex-post evaluation

Figure 1: Results of Experiments

$$Y(p) = \sum_{i=1}^{|P|} p_{a_i} * y_{a_i} \qquad \lambda = \frac{\sum_{i=1}^{|P|} y_{a_i}}{\sum_{k=1}^{|\overline{P}|} y_{a_k}} \tag{2}$$

where $p_{a_i}$ and $y_{a_i}$ are the percentage and the average yield of the *i-th* asset class in the portfolio, respectively. $Y(p)$ 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 $\lambda \ll 1$, it acts as a dumping factor). Through this strategy, all the *candidate solutions* are ranked according to the FCV score and the*top-k* solutions are returned to the advisor.

**(e) FCV + Greedy:** this combined strategy first uses the greedy algorithm to diversify the solutions, then exploits the FCV to rank the portfolios and obtain the *final solutions*.
**(3) Review and Retain:** in the REVIEW step human advisor and client can further discuss and modify the portfolio, before generating the *final solution* for the user. If the yield obtained by the newly recommended portfolio is acceptable, the solution is stored in the case base and can be used in the future as input to resolve similar cases.

## 2. EXPERIMENTAL EVALUATION

The performance of recommended portfolios, generated with different *revise strategies*, were compared to both a CF baseline and to the portfolios suggested by a human advisor. Next, an *ex-post evaluation* was performed by evaluating the real yield obtained by the portfolios after three months.

**Experimental Design.** Experiments were performed by exploiting a dataset of 1148 real (anonymous) users, which agreed their portfolios between June 2011 and June 2013. The dataset was provided by Objectway Financial Software and is available for download[6]. Each case in the *case base* was represented by adopting the previously introduced formalism. Feedback assessments were obtained by calculating the average yield of each portfolio from the agreement date to January 2014. To provide users with recommendations a *leave-one-out* design was adopted, that is to say, at each run, the case base was built by exploiting all the portfolios with the exception of the one agreed by the target user. Statistical differences were assessed by adopting a *paired t-test* on the average monthly yield of each portfolio, with $p < 0.05$.

**Discussion of the results.** Results of Experiment 1 are reported in Figure 1a, which shows the average yield obtained by each ranking strategy by picking the first $n$ portfolios ($n$=1, 5, 10) from a neighborhood of fixed size (50). The main outcome of this experiment is that FCV *significantly* outperforms cosine similarity as ranking strategy. Results show that through FCV it is possible to get a 0.3% yield per month, on average. The Greedy Diversification strategy got promising results as well: even if the obtained yield was worse, no statistical differences were noted for *n=5* and *n=10*. Finally, we compared the best configuration (FCV) to an adaptation of item-based CF to the financial domain and to the recommendations provided by a human advisor. As human recommendations we used the 1148 real portfolios stored in the case base, along with the yield they generated. Results (Figure 1b) show that our framework significantly outperforms baselines with $n$=5 and 10.

Next, in the *ex-post* evaluation of our framework we compared the (real) yield gained by the portfolios in the time interval between January and April 2014. In other terms, we simulated that the recommended portfolios were actually agreed in January 2014 and we analyzed the *real ex-post perfomance* of the recommendations generated by the framework. To this end, we first calculated FCV scores by using the historical yield of the asset classes up to January 2014. Next, we generated the recommendations and we calculated their yield *only* between January and April 2014. As shown in Figure 1c, this experiment confirmed the goodness of FCV and its (significant) improvement with respect to human baseline. The most interesting outcome of Experiment 2 was that the combination of the diversification technique with FCV further improves the performance of the proposed portfolios. This result suggests that a combined strategy which merges a ranking based on past performance, as FCV, with an algorithm leading to more diverse recommendations can make the framework more effective, since a human advisor is put in a position to base his investment proposal on the ground of many *diversified solutions*, which are more effective, especially when market fluctuations have to be tackled.

[6]http://bit.ly/financialRS_data_uniba

## 3. REFERENCES

[1] F. Lorenzi and F. Ricci. Case-based recommender systems: a unifying view. In *Intelligent Techniques for Web Personalization*, pages 89–113. Springer, 2005.
[2] M. Montaner, B. López, and J. L. de la Rosa. Improving case representation and case base maintenance in recommender agents. In *Advances in Case-Based Reasoning*, pages 234–248. Springer, 2002.
[3] B. Smyth and P. McClave. Similarity vs. diversity. In *Case-Based Reasoning Research and Development*, pages 347–361. Springer, 2001.