OLPS: A Toolbox for On-Line Portfolio Selection

Bin Li   BINLI.WHU@WHU.EDU.CN
Economics and Management School, Wuhan University
Wuhan, P.R. China 430072

Doyen Sahoo   DOYEN.2014@PHDIS.SMU.EDU.SG
School of Information Systems, Singapore Management University
Singapore 178902

Steven C. H. Hoi   CHHOI@SMU.EDU.SG
School of Information Systems, Singapore Management University
Singapore 178902

Editor:

Abstract

On-line portfolio selection is a practical financial engineering problem, which aims to sequentially allocate capital among a set of assets in order to maximize long-term return. In recent years, a variety of machine learning algorithms have been proposed to address this challenging problem, but no comprehensive open-source toolbox has been released for various reasons. This article presents the first open-source toolbox for "On-Line Portfolio Selection" (OLPS), which implements a collection of classical and state-of-the-art strategies powered by machine learning algorithms. We hope OLPS will facilitate the development of new learning methods and enable the performance benchmarking and comparisons of different strategies. OLPS is an open-source project released under Apache License (version 2.0), which is available at `https://github.com/OLPS/`.

Keywords: On-line portfolio selection, online learning, trading system, simulation.

1. Introduction

In recent years, machine learning has been applied to various applications in finance (Györfi et al., 2012), including On-line Portfolio Selection, which aims to sequentially allocate capital among a set of assets, such that the return of the investments can be maximized in the long run (Kelly, 1956). It has attracted increasing attentions from both academia and industry, and several machine learning algorithms have been proposed (Li and Hoi, 2014), including traditional algorithms (Cover, 1991; Helmbold et al., 1998; Agarwal et al., 2006; Borodin et al., 2004; Györfi et al., 2006, 2008), and recent state-of-the-art online-learning algorithms (Li et al., 2011, 2012; Li and Hoi, 2012; Li et al., 2013). Unlike other application domains in machine learning where various open-source packages are available, very few open-source toolkit [1] exists for on-line portfolio selection, primarily due to the confidential nature of financial industry. Consequently, it is difficult for researchers to evaluate new algorithms for comprehensive comparisons with existing ones.

This article introduces an open-source toolkit named "On-Line Portfolio Selection" (OLPS), which consists of a family of classical and state-of-the-art on-line portfolio selection algorithms. To

---

1. Some repositories contain public datasets, such as `http://www.cs.technion.ac.il/~rani/portfolios` and `http://www.szit.bme.hu/~oti/portfolio`, but no public code was released.

the best of our knowledge, OLPS is the first comprehensive open-source package for this problem, which includes various strategies, and a set of preprocessing, postprocessing and visualization tools in an integrated platform. The open-source nature of OLPS make it easy for a third party to develop new algorithms, and facilitate the comparisons with many in-built algorithms on real datasets.

2. Overview and Implementation

OLPS implements a framework for backtesting various algorithms of on-line portfolio selection in both Matlab [2] and Octave [3] (except that GUI is only available in Matlab) under Windows, Linux, and Mac OS. Figure 1 gives an overview of the OLPS toolkit with three main modules: (i) data preprocessing: it loads a dataset and initializes the environments, e.g., preparing log file handles, etc; (ii) algorithms: it calls some of the algorithms, and simulates the trading process with different strategies based on the data from the first module; and (iii) post-processing: it analyzes statistical significance of the results from the second module, e.g., some risk-adjusted returns.
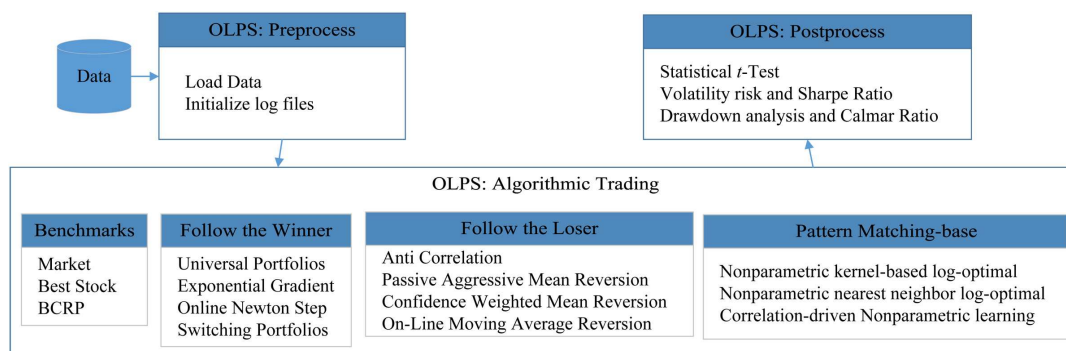

Figure 1: Structure of the OLPS toolbox.

2.1 Pre-processing

Datasets	Markets	Regions	Time Frames	# Periods	# Assets	File Names (.mat)
NYSE (O)	Stock	US	07/03/1962 - 12/31/1962	5651	36	`nyse-o`
NYSE (N)	Stock	US	01/01/1985 - 06/30/2010	6431	23	`nyse-n`
TSE	Stock	CA	01/04/1994 - 12/31/1998	1259	88	`tse`
SP500	Stock	US	01/02/1998 - 01/31/2003	1276	25	`sp500`
MSCI	Index	Global	04/01/2006 - 03/31/2010	1043	24	`msci`
DJIA	Stock	US	01/14/2001 - 01/14/2003	507	30	`djia`

Table 1: Summary of the six datasets from real markets.

The main functionality of this step is to initialize the trading environment. OLPS supports all types of datasets that Matlab/Octave accepts, and all existing datasets are in MAT-file [4]. The datasets

---

2. More details on Matlab are available at `http://www.mathworks.com`.

3. Octave can be freely downloaded from `https://www.gnu.org/software/octave/`

4. Full description about MAT file can be found at `http://www.mathworks.com`.

contain an $n \times m$ matrix of price relatives, where n denotes the number of trading periods, and m refers to the number of assets. It can be further adapted to incorporate real market data feeds [5], such that the toolkit can handle real time data and conduct paper trading or real trading [6].

Table 1 summarizes several representative public datasets included in the toolbox [7]. Among them, NYSE (O) is a benchmark dataset proposed by Cover (1991). The datasets TSE, SP500 and DJIA were first collected by Borodin et al. (2004). Györfi et al. (2006) and we collected the NYSE (N) dataset [8]. MSCI was first collected by us in our previous work (Li et al., 2012).

2.2 Algorithmic Trading

This step conducts simulation based on historical market data. In our framework, implementing a new strategy generally requires four files, i.e., an entry file, a run file, a kernel file and an expert file. The entry file extracts parameters and call corresponding run file. The run file simulates a whole trading process, and calls its kernel file to construct a portfolio for each period, which is used for rebalancing. The kernel file outputs the portfolio, while it facilitates the development of meta algorithms, which combines multiple experts' portfolios outputted by expert files. The expert file outputs one portfolio depending on input data and its parameters. In case of only one expert, the kernel file is not necessary and directly enters the expert file.

Categories	Algorithms	Entry Files	References
Benchmarks	Uniform BAH	`ubah.m`	Li and Hoi (2014)
	Best Stock	`best.m`	Li and Hoi (2014)
	Uniform CRP	`ucrp.m`	Li and Hoi (2014)
	BCRP	`bcrp.m`	Cover (1991)
Follow the Winner	UP	`up.m`	Cover (1991)
	EG	`eg.m`	Helmbold et al. (1998)
	ONS	`ons.m`	Agarwal et al. (2006)
	SP	`sp.m`	Singer (1997)
Follow the Loser	Anticor	`anticor/anticor_anticor.m`	Borodin et al. (2004)
	PAMR	`pamr/pamr_1/pamr_2.m`	Li et al. (2012)
	CWMR	`cwmr_stdev/cwmr_var.m`	Li et al. (2013)
	OLMAR	`olmar1/olmar2.m`	Li and Hoi (2012)
Pattern Matching based Algorithms	B^K	`bk.m`	Györfi et al. (2006)
	B^{NN}	`bnn.m`	Györfi et al. (2008)
	CORN	`corn/cornu/cornk.m`	Li et al. (2011)
Others	M0	`m0.m`	Borodin et al. (2000)
	T0	`t0.m`	Borodin et al. (2000)

Table 2: Summary of the implemented algorithms.

Table 2 summarizes the families of algorithms implemented in the OLPS toolbox, including
- Benchmarks;
- Follow the Winner approaches: make portfolio decisions following the assumption that the next price relatives (or experts for UP) will follow previous one;

---

5. For example, Interactive Brokers (`http://www.interactivebrokers.com`) provides free APIs.

6. Paper and real trading both require to implement order submission step, while back test does not.

7. More datasets and their details, including the components, can be found at the project web site.

8. The dataset before 2007 was collected by Gábor Gelencsér (`http://www.cs.bme.hu/~oti/portfolio`), and we collected the remaining data from 2007 to 2010 via Yahoo Finance.

- Follow the Loser approaches: make portfolio decisions by assuming that next price relatives will revert to previous trends;
- Pattern Matching-based approaches: locate a set containing indices of similar historical patterns, and make optimal portfolio decisions thereafter.

2.3 Post-processing

After the algorithmic trading simulation, the last step is to post-process the results for performance analysis by summarizing the following performance metrics:

- Cumulative Return: the most widely used performance metric in related studies;
- Volatility and Sharpe Ratio: measures volatility risk and related risk-adjusted return;
- Drawdown and Calmar Ratio: measures downside risk and risk-adjusted return;
- T-test statistics (Grinold and Kahn, 1999): tests if a strategy's return is significantly different from the market.

3. Usage Example

We provide two interfaces to run simulations in the toolbox. One is the Graphical User Interface (GUI) (available in Matlab) and the other is a Command Line Interface (CLI). While GUI is straightforward to use, the following example illustrates its CLI usage. Suppose we are going to simulate PAMR algorithm on the NYSE (O) dataset. The commands can be listed as follows:

```
1: >> opts.quiet_mode=1; opts.display_interval=500; opts.progress=0;
2: >> opts.analyze_mode=1; opts.log_mode=1; opts.mat_mode=1;
3: >> manager('pamr', 'nyse-o', {0.5, 0}, opts);
```

Line 1 and 2 set the control variables, which can be used multiple times [9]. Line 3 calls the manager function to call "pamr" strategy on "nyse_o" with the parameters equaling "{0.5, 0}". During the simulation, the algorithm will output information periodically. After the simulation, the algorithm will analyze the returns and output the cumulative return, risk-adjusted returns and statistical test statistics. Corresponding running details and results will be automated stored in the Log folder.

4. Summary

This article presented OLPS — an open-source On-Line Portfolio Selection toolbox to facilitate the related research in machine learning and computational finance. This is the first open-source project in this area, which not only facilitates researchers to develop new strategies, but also allows them to easily benchmark their performance with existing strategies. Besides, the toolbox supports a large collection of classical and state-of-the-art on-line portfolio selection strategies, provides three different performance metrics: cumulative return, volatility risk-adjusted return, and drawdown risk-adjusted return. The toolbox offers a user-friendly GUI in Matlab and also can run in the command-line interface for both Matlab and Octave. We hope the open-source nature of the software (available under Apache License version 2.0) would encourage researchers to extend the toolkit and share their algorithms through the OLPS platform. More documentations and resources, including a comprehensive survey, can be found in our project website: `http://OLPS.stevenhoi.org/`.

---

9. Further details of these control variables can be found in the manual.

References

Amit Agarwal, Elad Hazan, Satyen Kale, and Robert E. Schapire. Algorithms for portfolio management based on the newton method. In *Proceedings of International Conference on Machine Learning*, 2006.

Allan Borodin, Ran El-Yaniv, and Vincent Gogan. On the competitive theory and practice of portfolio selection (extended abstract). In *Proceedings of the Latin American Symposium on Theoretical Informatics*, 2000.

Allan Borodin, Ran El-Yaniv, and Vincent Gogan. Can we learn to beat the best stock. *Journal of Artificial Intelligence Research*, 21:579–594, 2004.

Thomas M. Cover. Universal portfolios. *Mathematical Finance*, 1(1):1–29, 1991.

Richard Grinold and Ronald Kahn. *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk*. New York : McGraw-Hill, New York, 1999.

László Györfi, Gábor Lugosi, and Frederic Udina. Nonparametric kernel-based sequential investment strategies. *Mathematical Finance*, 16(2):337–357, 2006.

László Györfi, Frederic Udina, and Harro Walk. Nonparametric nearest neighbor based empirical portfolio selection strategies. *Statistics and Decisions*, 26(2):145–157, 2008.

László Györfi, Gy. Ottucsák, and Harro Walk. *Machine Learning for Financial Engineering*. Singapore : World Scientific, 2012.

David P. Helmbold, Robert E. Schapire, Yoram Singer, and Manfred K. Warmuth. On-line portfolio selection using multiplicative updates. *Mathematical Finance*, 8(4):325–347, 1998.

Jr. Kelly, J. A new interpretation of information rate. *Bell Systems Technical Journal*, 35:917–926, 1956.

Bin Li and Steven C. H. Hoi. On-line portfolio selection with moving average reversion. In *Proceedings of the International Conference on Machine Learning*, 2012.

Bin Li and Steven C. H. Hoi. Online portfolio selection: A survey. *ACM Computing Surveys*, 46 (3):35, 2014.

Bin Li, Steven C.H. Hoi, and Vivekanand Gopalkrishnan. Corn: Correlation-driven nonparametric learning approach for portfolio selection. *ACM Transactions on Intelligent Systems and Technology*, 2(3):21:1–21:29, 2011.

Bin Li, Peilin Zhao, Steven Hoi, and Vivekanand Gopalkrishnan. Pamr: Passive aggressive mean reversion strategy for portfolio selection. *Machine Learning*, 87(2):221–258, 2012.

Bin Li, Steven C.H. Hoi, Peilin Zhao, and Viveknand Gopalkrishnan. Confidence weighted mean reversion strategy for online portfolio selection. *ACM Transactions on Knowledge Discovery from Data*, 7(1):4:1 – 4:38, 2013.

Yoram Singer. Switching portfolios. *International Journal of Neural Systems*, 8(4):488–495, 1997.