Optimal Asset Allocation using Adaptive Dynamic Programming

Neuneier. Ralph, In *Advances in Neural Information Processing Systems.* 1996.

Enhancing Q-Learning for Optimal Asset Allocation

Neuneier. Ralph, In *Advances in Neural Information Processing Systems.* 1998.

強化学習勉強会　第 50 回

2017/3/8　西村 直樹

本論文を選んだ動機

- 実務においてリスクを考慮した上での意思決定が求められる状況は多く存在する
- リスクを考慮した意思決定問題の代表的なものとしてポートフォリオ最適化問題があり、資産価値を最大化したい、というのは時代を超えて関心を集める問題である
- 本論文は、ポートフォリオ最適化問題をMDPとして表現し、動的計画法のアプローチで取り組んだ論文であり、応用としても興味深いと感じたため紹介する

論文の貢献

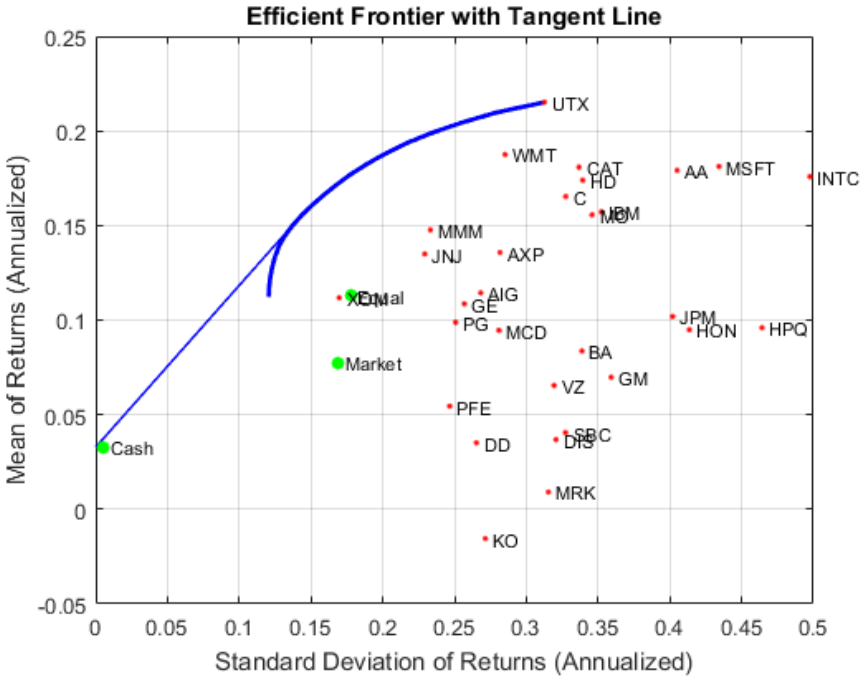
論文の貢献として以下のことが挙げられている

1. いくつかの仮定のもとでポートフォリオ最適化問題をMDPとしてモデル化した
2. 取引コストやリスク性向といった古典的なポートフォリオ最適化における条件についても考慮した
3. 人工データとドイツ株式指数(DAX)に対して他モデルによる戦略と性能を比較する数値実験を行った

目次

ポートフォリオ最適化問題

- 投資家が求める期待収益率を達成した上で、リスクを最小にするような各投資対象に対する投資比率を決定する問題
- Mathworks： Portfolio Optimization Examples より
  - 各資産収益率の平均と分散、効率的フロンティアの図



https://jp.mathworks.com/help/finance/portfolio-optimization-examples.html?requestedDomain=www.mathworks.com

古典的なポートフォリオ最適化問題

- 以下の 2 ステップでポートフォリオを構築
  1. それぞれの資産に対する期待収益を予測
  2. 収益率予測と投資家のリスク性向に応じてポートフォリオを構築
- 平均・分散モデル [Markowitz 1952]
  - 要求期待収益率以上で分散が最小となる投資比率を決定

$$\textbf{Minimize} \quad \sum_{j=1}^{n}\sum_{k=1}^{n}\sigma_{jk}x_j x_k$$

$$\textbf{subject to} \quad \overline{r}_p \geq r_E$$

$$\sum_{j=1}^{n} x_j = 1$$

$$x_j \geq 0, \ (j = 1, \ldots, n)$$

$$\boldsymbol{x} \in X$$

■ 決定変数

x_j: 資産(証券) j の投資比率。$\boldsymbol{x} = (x_1, \ldots, x_n)^T$

■ パラメータ

n: 資産(証券)数

T: 期間数もしくはシナリオ数

r_{jt}: 事象 t における資産 j の収益率

r_E: 投資家の要求期待収益率

$\overline{r}_j$: 資産(証券) j の期待収益率

σ_{jk}: 資産(証券) j と資産(証券) k の共分散

ポートフォリオの期待収益率 $\overline{r}_p$

$$\overline{r}_p = \sum_{j=1}^{n}\overline{r}_j x_j$$

引用：平均分散モデル (1) http://www.ae.keio.ac.jp/lab/soc/hibiki/portfolio/mv_1.pdf

古典的なポートフォリオ最適化における課題

- 論文中では、以下のことが古典的ポートフォリオ最適化において問題を複雑にしていると述べられている
  - ポートフォリオのリバランスのために時間ステップごとに計算を行わなければならない
  - リバランスにおける取引コストを考慮しなければならない

目次

問題の単純化のための仮定

- 以下の単純化のための仮定により、MDPとポートフォリオ最適化の関係を明確にすることができる
    1. ドイツマルクを基にする投資家にとって、USドルを唯一の投資可能な資産とする
    2. 投資家は小さく取引によって市場に影響を与えない
    3. 投資家は全ての資産を投機のために利用する
    4. 投資家は無限の時間軸の各時間ステップ取引可能である
- 1, 3 の仮定については [Neuneier 1998] で緩和する方法について考察されている

ポートフォリオ最適化問題のMDPとしての表現

- マルコフ決定過程（MDP）
  - $\$_t$ ：金融市場（金利、株式指標など）
  - K_t ：t 時点の資産配分
  - $x_t = (\$_t, K_t)'$ ：金融市場と資産配分の状態
  - $a_t = \mu(x_t)$ ：状態における方策 μ における行動
  - $p(x_{t+1}|x_t)$ ：状態遷移確率
  - $r(x_t, a_t, \$_{t+1})$ ：状態において行動をとったときの報酬

ポートフォリオ最適化問題のベルマン最適方程式

- 最適状態価値関数

$$V^*(x_t) = \max_a \left[\sum_{x_{t+1}} p(x_{t+1}|x_t, a) r_t + \gamma \sum_{x_{t+1}} p(x_{t+1}|x_t, a) V^*(x_{t+1}) \right]$$

- 最適行動価値関数

$$Q^*(x_t, a_t) := \sum_{x_{t+1}} p(x_{t+1}|x_t, a_t) r_t + \gamma \sum_{x_{t+1}} p(x_{t+1}|x_t, a_t) \max_{a \in A} (Q^*(x_{t+1}, a))$$

ポートフォリオ最適化問題の Q 学習

- Q 学習

$$\text{QL: } Q_{k+1}(x_t, a_t) = (1 - \eta_k) Q_k(x_t, a_t) + \eta_k (r_t + \gamma \max_{a \in A}(Q_k(x_{t+1}, a)))$$

学習率　　割引率

- 行動価値関数の Neural Network (NN) での関数近似
  - 金融市場と資産配分の状態空間は連続で広い

$$d_t := r(x_t, a_t, x_{t+1}) + \gamma \max_{a \in A} Q(x_{t+1}; w_k^a) - Q(x_t; w_k^{a_t}) ,$$

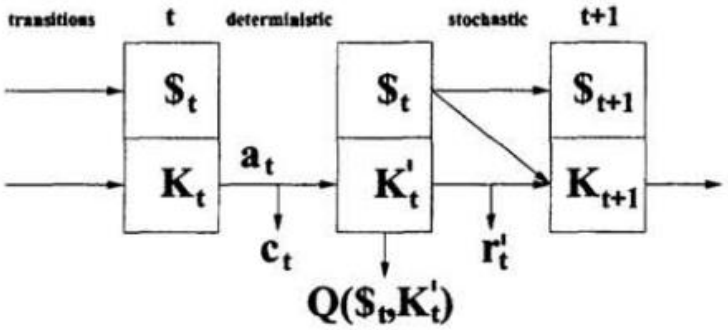
$$\text{NN-QL: } w_{k+1}^{a_t} = w_k^{a_t} + \eta_k d_t \nabla Q(x_t; w_k^{a_t}) .$$

目次

取引コストを考慮したベルマン方程式

- 価値関数は決定的な報酬 (取引コスト) と確率的な報酬 (市場変化による報酬) に分解できる

$$V^*(\$, K) = \max_{a_0,\cdots} E\left[\sum_{t=0}^{\infty} \gamma^t (c(K_t, a_t) + r'(\$_t, K'_t, \$_{t+1})) \,\middle|\, \begin{matrix} \$_0 = \$ \\ K_0 = K \end{matrix}\right]$$



- 取引コストを考慮した行動価値関数

$$\begin{aligned} Q^*(\$_t, K'_t) &:= E\left[r'(\$_t, K'_t, \$_{t+1}) + \gamma V^*(\$_{t+1}, K_{t+1})\right] \\ &= E[r'_t + \gamma \max_{a_{t+1}}[c_{t+1} + Q^*(\$_{t+1}, K'_{t+1})]] \end{aligned}$$

取引コストを考慮したポートフォリオ最適化問題の Q 学習

- 取引コストを考慮したベルマン最適方程式（ t は省略）

$$V^*(\$, K) = \max_a [c(K, a) + Q^*(\$, K')],$$

$$\mu^*(\$, K) = \arg\max_a [c(K, a) + Q^*(\$, K')]$$

- TD誤差

$$d_t := r'(\$_t, K'_t, \$_{t+1}) + \gamma \max_a [c(K_{t+1}, a) + Q^{(k)}(\$_{t+1}, K'_{t+1})] - Q^{(k)}(\$_t, K'_t)$$

- 取引コストを考慮した行動価値関数の近似

$$\text{QLU:} \quad Q^{(k+1)}(\$_t, K'_t) = Q^{(k)}(\$_t, K'_t) + \eta_k d_t \,,$$

$$\text{NN-QLU:} \quad w^{(k+1)} = w^{(k)} + \eta_k d_t \nabla Q(\$, K'; w^{(k)})$$

最適意思決定のためのQLUのアルゴリズム

1. データセットからランダムに $\$_t, \$_{t+1}$ を抽出しランダムに投資配分 K'_t を決定
2. 任意のとりうる行動 a について $r'_t, c(K_{t+1}, a), Q^{(k)}(\$_{t+1}, K'_{t+1})$ をそれぞれ計算
3. TD誤差 d_t を計算
4. $Q^{(k+1)}(\$_t, K'_t)$ または $Q(\$_t, K'_t; w^{(k+1)})$ を計算
5. Q値が収束したら終了、そうでなければ 1. へ戻る

- Q 学習と同じ Q 値に収束する
- 利点はいくつの資産が存在したとしても 1 つの価値関数しか必要なく、とりうる行動の粒度のみがパラメータとなる点である
- 1. により、十分な状態と行動の探索が保証される

リスク調整MDP

- 有限状態空間において定常方策 $\mu(x)$ が与えられたとし、
  価値関数とその分散を以下のように定義する

$$V^{\mu}(x) = E\left[\sum_{t=0}^{\infty}\gamma^{t} r(x_t, \mu_t, x_{t+1}) \middle| x_0 = x\right],$$

$$\sigma^2(V^{\mu}(x)) = E\left[\left(\sum_{t=0}^{\infty}\gamma^{t} r(x_t, \mu_t, x_{t+1}) - V^{\mu}(x)\right)^2 \middle| x_0 = x\right]$$

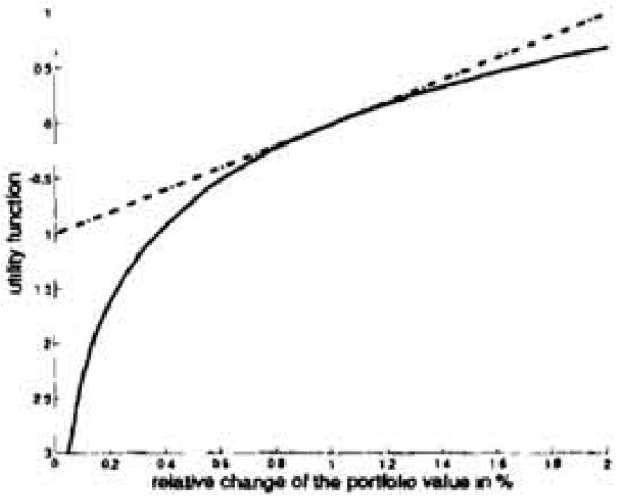
- リスク調整MDPにおける最適戦略は以下のように定式化できる

$$\mu^*(x;\lambda) = \arg\max_{\mu}[V^{\mu}(x) - \lambda\,\sigma^2(V^{\mu}(x))] \quad \text{for } \lambda \geq 0$$

- λ の値を調整することでリスク性向に応じたポートフォリオが
  構築できる（正則化パラメータのようなイメージ）

非線形効用関数

- $\sigma^2(V^\mu(x))$ は再帰的ベルマン方程式として記述できないため一般的に計算できない
- リスクを表す指標としてシャープレシオ $\overline{r}/\sigma(r)$ が用いられることが多いが upside volatility に対してもペナルティがかかってしまう
- そこで $r = log(\text{new portfolio value / old portfolio value})$ のように増加による効用増より損失による効用減が大きくなるような非線形の効用関数が用いられることもある



参考：その他のリスク指標

- これからの強化学習 (2016) 2.6 節 リスク考慮型強化学習 にて以下の 3 タイプのリスク考慮型強化学習法が紹介されている

1. ある種の最悪のケース評価 [Heger 1994]

$$Q(i,a) := \max\left[Q(i,a),\ r+\gamma\cdot\min_{b\in A(j)}Q(j,b)\right]$$

2. 効用関数や時間差分誤差の非線形化

   [Mihatsch and Neuneier 2002]

$$\mathcal{X}^{\kappa}: x \mapsto \begin{cases}(1-\kappa)x & \text{if } x>0,\\ (1+\kappa)x & \text{otherwise.}\end{cases} \quad \kappa\in(-1,1)$$

   TD誤差　　リスク選好性

3. リターン以外のリスク指標を導入するアプローチ

   [Geibel and Wysotzki 2005]

$$V_{\xi}^{\pi}(x) = \xi V^{\pi}(x) - \rho^{\pi}(x)$$
$$Q_{\xi}^{\pi}(x,u) = \xi Q^{\pi}(x,u) - \bar{Q}^{\pi}(x,u)$$

エラー状態に到達する確率の関数

目次

- ポートフォリオ最適化問題とは
- ポートフォリオ最適化問題の MDP としての表現
- ポートフォリオ最適化問題のための Q 学習の拡張
- **数値実験**
- まとめ

数値実験：人工為替相場 [Neuneier 1996]

- 問題設定：DM (ドイツマルク)か US-$ のどちらかに投資をする
- x_t：US-$ 1 単位に対する DM の為替相場 [1,2]
- p_{ij}: 為替相場の推移確率 (増加する確率が高いが、値が高いとき
  とても低い値に落ちる確率が高い)
- u_t：投資家の決定 (DM に投資するか US-$ に投資するか)
- c_t：ポートフォリオの富 (負値は US-$ で絶対値の DM 分の価値)
- $\xi = 0.1 + |c/100|$: 取引コスト（左項：固定費、右項：変動費）
- 即時報酬：下表に基づき計算される

$r_t(x_t, c_t, x_{t+1}, u_t)$	$u_t = \text{DM}$	$u_t = \text{US-\$}$
$c_t \in \text{DM}$	0	$r_t = (x_{t+1}/x_t)(c_t - \xi) - c_t$
$c_t \in \text{US-\$}$	0	$r_t = (x_{t+1}/x_t - 1)c_t$

数値実験：人工為替相場 [Neuneier 1996]

■ Q学習

- 即時報酬と状態を用いて Q 学習を行う
- 学習率は 0.1 とする
- Q 値は0で初期化し、2000時間ステップを1期として4期で収束した

$$
\begin{aligned}
Q(i,u(i)) &= (1-\eta)Q(i,u(i)) + \eta(r(i,j,u(i)) + \gamma \max_{u(j)} Q(j,u(j))) \\
Q(k,v) &= Q(k,v), \text{ for all } k \neq i \text{ and } v \neq u(i)
\end{aligned}
$$

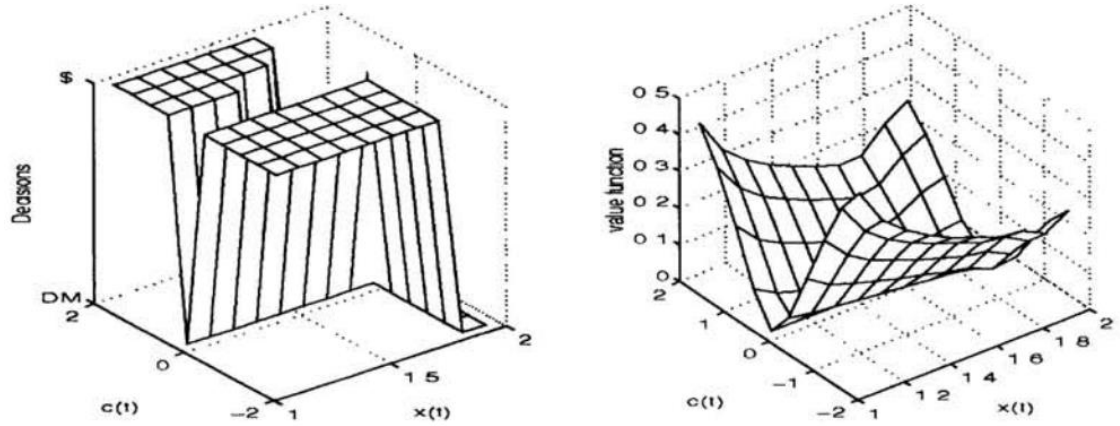
■ Q学習で得られた解の評価

- 人工データ作成の際の状態遷移確率を既知として価値関数を推定
- 500時間ステップで収束

$$
T_i(V) = \max_{u(i) \in U(i)} (R(i,u(i)) + \gamma \sum_{j \in S} p_{ij}^{\pi} V_j).
$$

数値実験：人工為替相場 [Neuneier 1996]

- 学習された Q 関数に基づく意思決定と価値関数の比較

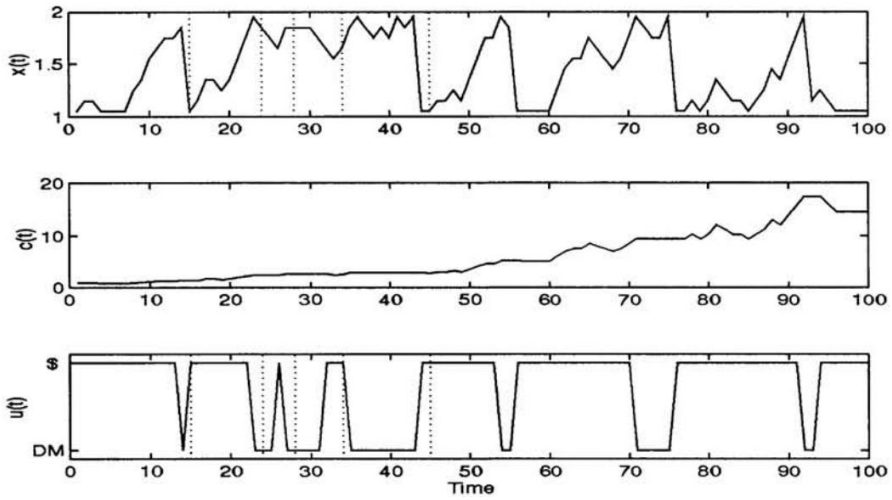


- 学習されたモデルの考察
  - US-$を保有している場合はリスクが高くなるまで保持
  - DMを保有している場合はリスクが低い場合にUS-$に投資

  (為替相場の上昇による期待利得が取引コストに見合わない)

数値実験：人工為替相場 [Neuneier 1996]

- 100日間の為替相場 x_t 累積資産 c_t 投資決定 u_t



- 結果の考察
  - 最初はすぐにUS-$へ、13時間ステップまで保持
  - 35～45時間ステップの振動している部分ではリスクが高いとしDMへ
  - 24～28時間ステップでは一時的にUS-$に変えすぐDMへ戻す
  - 累積資産は全体を通して増加しており、戦略が成功している

数値実験：ドイツ株式指数 (DAX) [Neuneier 1996]

- 問題設定：DMかDAXに連動する投資信託どちらかに投資をする
- 比較戦略：MLP [Dichtl 1995]
  （NNにより予測、投資信託の収益率が正のときに投資）
- 入力データ：DAXと11のマーケットに影響を与える変数
- 価値関数近似：NN（8つの隠れユニット）
- 学習期間：1986/1/2 – 1992/12/31
- テスト期間：1993/1/2 – 1995/10/18
- 取引コスト：固定費 0.001単位、変動費 0.4 %
- 学習率：0.001

数値実験：ドイツ株式指数 (DAX) [Neuneier 1996]

- QL戦略はMLP戦略と比較し、学習期間は 80 %、テスト期間は 25 %最終時点で収益が増加

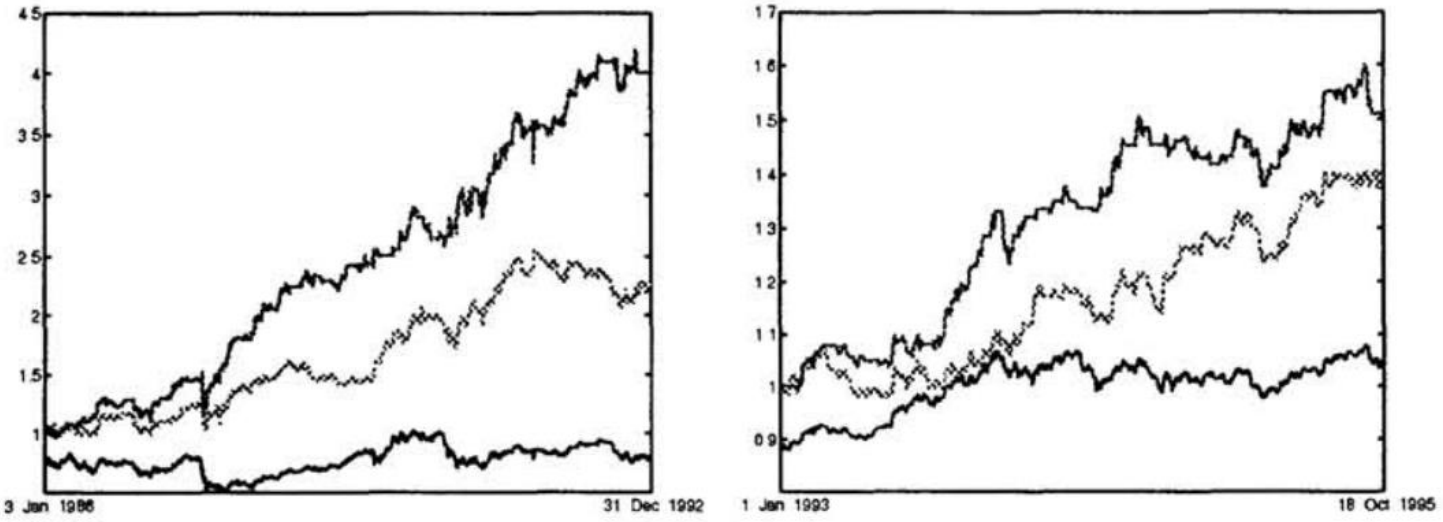


Figure 3: The development of a reinvested capital on the training (left) and test set (right). The lines from top to bottom: QL-strategy, MLP-strategy, scaled DAX.

数値実験：ドイツ株式指数 (DAX) [Neuneier 1996]

- 各方策の投資時間ステップ数とポートフォリオ組換え回数

		DAX investments		position changes	
	Data	MLP Policy	QL-Policy	MLP Policy	QL-Policy
Training set	1825	1020	1005	904	284
Test set	729	434	395	344	115

- 結果の考察
  - QL方策はMLP方策と同程度の時間ステップ数でDAXに対して投資しているがポートフォリオ組換え回数が少ないため取引コストが抑制できた

数値実験：ドイツ株式指数 (DAX) [Neuneier 1998]

- 問題設定：DMかDAXに連動する投資信託どちらかに投資をする
- 比較戦略： Neuro-Fuzzy [Neuneier 1996]
  （予測モデルによる投資信託の収益率が正のとき投資）
  Naive Prediction (過去のDAX収益率が正のとき投資)
  Buy&Hold (最初に投資し最終時点で売る)
- 入力データ：DAXと11のマーケットに影響を与える変数
- 学習期間：1986/1/2 – 1994/12/31
- テスト期間：1993/1/2 – 1996/8/1
- 取引コスト：0.2 %
- 学習率：10000時間ステップについて $\eta_0 = 0.05$、$\eta_k = \eta_0 \cdot 0.999^k$

数値実験：ドイツ株式指数 (DAX) [Neuneier 1998]

- 学習(テスト)の各戦略の利益、投資時間ステップ数、ポートフォリオ組換え回数

strategy	profit	investments in DAX	position changes
NN-QLU	1.60 (3.74)	70 (73)%	30 (29)%
Neuro-Fuzzy	1.35 (1.98)	53 (53)%	50 (52)%
Naive Prediction	0.80 (1.06)	51 (51)%	51 (48)%
Buy&Hold	1.21 (1.46)	100 (100)%	0 (0)%

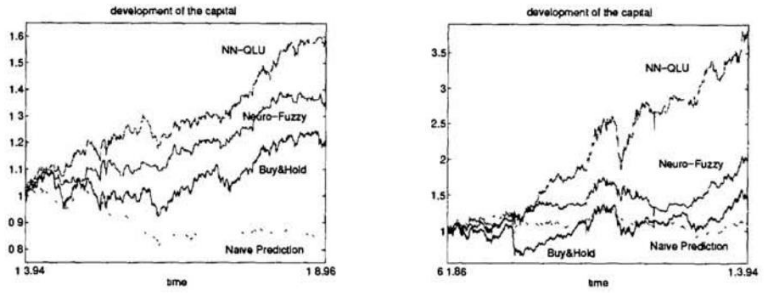


Figure 1: Comparison of the development of the capital for the test set (left) and the training set (right). The NN-QLU strategy clearly beats all the benchmarks.

- 結果の考察
  - 組換え回数が少なく取引コストが抑制でき、2位のNeuro-Fuzzyに18.5%(89%)の差をつける運用成績を上げている

数値実験：ドイツ株式指数 (DAX) [Neuneier 1998]

- 学習の振る舞い

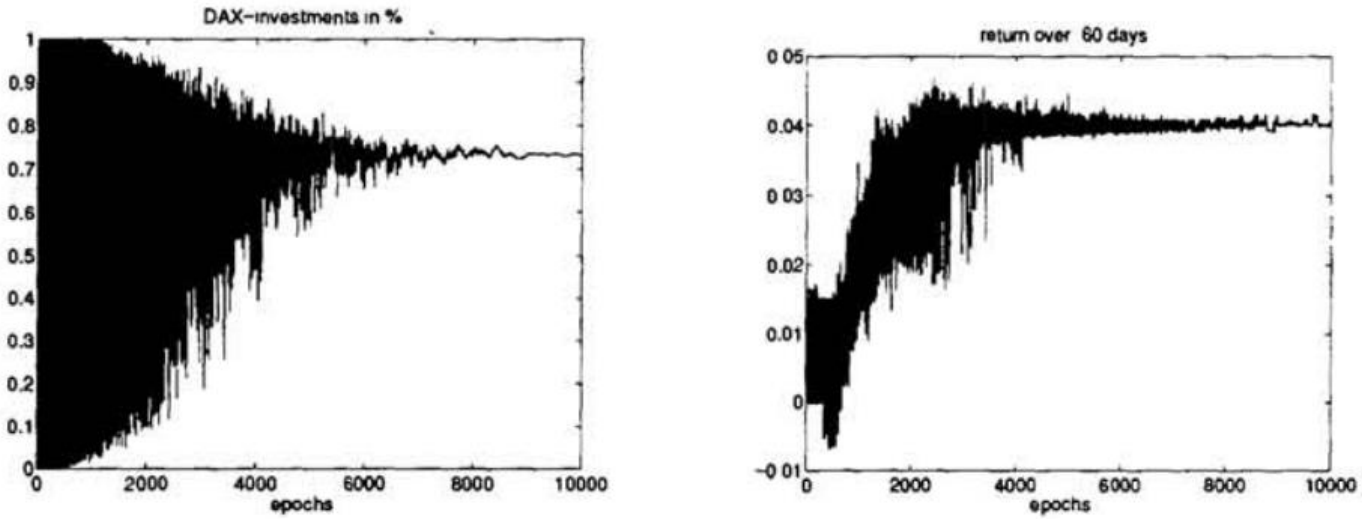


Figure 2: Training course: percentage of DAX investments (left), profitability measured as the average return over 60 days on the training set (right).

- 学習の初期では、全て投資するか全く投資しないかの方策であったが後期になると安定し収益も大きくなる

目次

まとめ

- 本論文ではポートフォリオ最適化問題をMDPとしてモデル化した
- ポートフォリオ最適化問題に特有の取引コスト、リスクの概念についても Q 学習において考慮した
- いくつかのベンチマークの戦略と比較し、提案手法である動的計画法によるアプローチの有効性が示された
- 方策のリスクを明示的に計算することが当時の本研究の Future work とされている