

Title	プログラム枚挙と統計的学習の組合せによる機械学習手法の提案と実装
Sub Title	A proposal of a machine learning method combining enumeration and statistical learning
Author	櫻井, 彰人(Sakurai, Akito)
Publisher	
Publication year	2016
Jtitle	科学研究費補助金研究成果報告書 (2015.)
JaLC DOI	
Abstract	<p>外国為替・株等の金融市場は極めて効率的であり、外国為替交換レートや株価・株価指標値等の短期金融時系列は、外部情報に対する瞬時的な応答を除けば、独立増分を持つ確率過程で近似できると考えられる。しかし実際には、当該応答の影響は暫時残り、増分は事前の動作・状態に依存している。本研究では、事前状態を、トレンドという人間の直観に合う形での記述を試み、また非線形関数で暗黙的に記述し、予測精度の向上を目指した。前者ではランダム性が強く出、予測精度の改善はなかったが、後者では十分な改善効果が得られた。</p> <p>Financial markets where foreign currency and stocks are exchanged are very efficient and short term time series of values of exchange rates, prices, and indexes are well approximated by stochastic processes with independent increments except for sudden changes caused by outer world events. But in reality, the changes remain and the increments are dependent on previous states/movements. In this research, so-called trends as human experts recognize are tried, and implicit non-linear representations are also tried, and used to describe and predict the time series. We showed that the former results in strong random sequences and does not contribute to improvements of prediction but the latter does.</p>
Notes	研究種目：基盤研究(C)(一般) 研究期間：2013～2015 課題番号：25330266 研究分野：機械学習
Genre	Research Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KAKEN_25330266seika

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the Keio Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

科学研究費助成事業 研究成果報告書

平成 28 年 5 月 22 日現在

機関番号：32612

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25330266

研究課題名(和文) プログラム枚挙と統計的学習の組合せによる機械学習手法の提案と実装

研究課題名(英文) A proposal of a machine learning method combining enumeration and statistical learning

研究代表者

櫻井 彰人 (Sakurai, Akito)

慶應義塾大学・理工学部・教授

研究者番号：00303339

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：外国為替・株等の金融市場は極めて効率的であり、外国為替交換レートや株価・株価指標値等の短期金融時系列は、外部情報に対する瞬時的な応答を除けば、独立増分を持つ確率過程で近似できると考えられる。しかし実際には、当該応答の影響は暫時残り、増分は事前の動作・状態に依存している。本研究では、事前状態を、トレンドという人間の直観に合う形での記述を試み、また非線形関数で暗黙的に記述し、予測精度の向上を目指した。前者ではランダム性が強く出、予測精度の改善はなかったが、後者では十分な改善効果が得られた。

研究成果の概要(英文)：Financial markets where foreign currency and stocks are exchanged are very efficient and short term time series of values of exchange rates, prices, and indexes are well approximated by stochastic processes with independent increments except for sudden changes caused by outer world events. But in reality, the changes remain and the increments are dependent on previous states/movements. In this research, so-called trends as human experts recognize are tried, and implicit non-linear representations are also tried, and used to describe and predict the time series. We showed that the former results in strong random sequences and does not contribute to improvements of prediction but the latter does.

研究分野：機械学習

キーワード：金融時系列予測 機械学習

1. 研究開始当初の背景

金融時系列の分析では、しばしば、市場変化の検出を目指す。数日や数か月にわたる市場の動きが、ある一瞬で非常に異なる動きに変わることがあり、市場参加者は、こうした変化を素早く検出し、投資方針をそれに合うよう変更し損失を防ぎたいと考えているからである。しかし、それは容易ではなく、従って多くの研究が行われている (J.D. Hamilton, Regime Switching Models.)。

こうした研究の多くは volatility の予測を目指し、価格の予測を目指してはいない。価格の変動はほぼランダムと考えられているからである。一方、実務家は、テクニカル分析によるトレンド予測は(ある程度)可能と考えている (C.-H. Park & S.H. Irwin, AgMAS Project Research Report 2004-04)。なお、直接に価格の予測を目指す以外に、自動売買を行う取引規則を作り、実データを用いた仮想的な取引で収益をあげうることを示す、予測可能性を間接的に示す方法もある。

近年ではアルゴリズム取引が頻繁に行われるようになった。専門的かつ大口の取引業者においては、高速取引の利点を生かすためにもアルゴリズム取引は必須である。一方、一般・小口取引者に対してアルゴリズム取引を許す仲介業者が現れている。通信速度等の制約があるため、高速取引とは言い難いが、アルゴリズムに従った自動的な取引ではある。

アルゴリズム取引においては、必ずしも従来のようなトレンド認識やそれを示唆するテクニカル指標を用いて売買タイミングを決める必要はない。機械的に過去のデータに基づき、予測を行えばよい。しかし、予測関数を作成するにあたり、従来の経験の塊であるテクニカル指標や取引業者が認識するトレンドに基づく予測も行われていることもまた事実である。

2. 研究の目的

金融時系列には状態の非連続な変化 regime switching がある。金融商品の市場価格の値動きにおけるトレンドの変化・転換とほぼ同じである。こうした時系列にはフラクタル性・再帰性があり、トレンドの中に、より短期のトレンドとトレンドの変化が観測される。一方、周期性は弱い。

このトレンドを補足する方法を開発し、それにより、市場価格の値動きの予測が可能である (最良気配値による仮想売買で収支が正になる) ことを示す

また、アルゴリズム取引に対抗することを想定した、トレンドの認識、結果としての予測、売買タイミングの最適化を検討する。

3. 研究の方法

本研究では、実データに存在する、トレンドのこの階層的・再帰的構造が簡易に表現でき、その学習を目指すアルゴリズムが簡易に

記述できる枠組み (プログラミング言語) を提案し、ツール (実行環境) を作成し、このツールを金融時系列の予測に適用する。

機械的に、各種のテクニカル指標や他の金融商品の価格を変数とする非線形関数を用いた予測手法の予測可能性を検討する。具体的な予測関数は、過去データを用いて学習的に取得する。学習された予測関数の予測性のがよいときには、その構造を調べ (予測可能ということは、何らかの経済状況の現時点での値が、過去から将来にわたって一定していることを意味し、それこそが広義でのトレンドと解釈できる故) その意味を検討する。

4. 研究成果

いくつかのトレンドの記述方法とそれによる予測手法を試みた。トレンド (より正確には、トレンドを一般化した値動きのパターン) の記述は可能であり、生成規則によるその一般化が可能であり、それに基づき予測を行うことは可能であるが、時系列予測の可能性を主張することはできなかった。

本研究においては、トレンドにというような明確な、すなわち、人間がみて認識できるようなパターン、すなわち、何らかの意味において単純なパターンの再生に基づく予測を試みているわけであるが、これを一般化し、単純な関数形による時系列予測の検討も行った。これは、例えば、いわゆるテクニカル指標を用いた予測に通じるわけであるが、本研究においては、それにはとらわれず、ある単純な関数群によって単純に表現される関数による予測も検討した。こちらからは意外な結果が得られた。

以下、トレンドに関する方法、単純な規則による方法について記し、本研究による大量の実験の結果から推測される金融時系列の特殊性について述べる。

金融時系列のスケールに依存しない値動きパターントレンドを表現するために、値動きを記号化する。説明の都合上単純な場合のみを記す。また数値としては、外国為替証拠金取引 (FX と呼ばれる) の内、ドル円為替レートに係る数値を用いる。

ドル円為替レートのスプレッドは、国内一般顧客向け業者では 0.3 銭固定が普通になってきているが、ICAP のような大口顧客を相手とする欧米大手では、呼値 0.5 銭単位であり、中央値 1.5 銭、平均 2 銭弱である。なお、1 分当りの最良買い気配値の値動きの幅の平均値は 0.5 銭前後であり、10 分当りのその標準偏差は 5.3 銭前後である (いずれも 2015 年)。

そこで、10 分足 (実際には他の足も検討) の return を考え、5 銭 (実際には他の長さも検討) を単位として離散化する。上昇、下降、停滞の 3 終端記号 (または述語) を用意し、これを単語とする単語列を作り、その規則性とそれに基づく予測可能性を検討した。

最初に n-gram 統計、それに基づく予測を

検討した。ほぼランダムであり予測に用いることはできなかった。粗視化したり、それとの混合を試みたりしたが、予測に用いても予測精度の向上は測れなかった。

より大きなパターンを発見するために、上記の単語列を言語と考える言語モデルを用いた。対象単語列を終端記号列とする文法を作成した。(正しい意味での)正規分布や文脈自由文法と異なり、属性値を非終端記号間で受け渡せる属性文法の形態をとる。属性値としては、ドル円為替レート、その変化、当該非終端記号が覆う時系列の継続時間等を試みた。

正規文法(但し、属性伝播は行う)による記述では、直観的には(何回かの上昇・下降を繰り返した後、すなわち、これを条件とした)現在のトレンドを記述することになる。従って予測は、(何回かの上昇・下降を条件とした)現在のトレンドの継続/終了/反転を予測することになる。

このパターンは非常にランダム性が強いために、予測を行うことは困難であることが分かった。粗視化と組み合わせても改善は得られなかった。

文脈自由文法(属性伝播を伴う)による記述により、直観的には山または谷の形を持つ値動きを値動きパターンの記述に加えることを試みた。単峰の山の形は、一連の上昇と下降がこの順でほぼ同じ長さ起こると記述する。多峰の山の形は、単峰の山が連なり、さらにその左右に同じ長さの上昇と下降があるとして記述でき、さらにその再帰的な形として記述することができる。谷形のパターンは、山形のそれにおいて上昇と下降を逆に用いて定義することができる。しかし、山形と谷形を同時に用いることは、値動きパターンの解釈(構文解析)結果が曖昧となる(複数の解釈が可能となる)ことから、できない。

構文解析には Ruby 版の yacc である racc を用いた。yacc/racc は LALR パーサーであり、言語を設計するわけではない(解析できるように言語を設計するわけではない)ため、実際、パターンのある特徴の記述を図ると reduce/reduce conflict が発生してしまう場合がある。現状では reduce/reduce conflict があっても動作するよう racc は構成されているが、希望の動作を行わせるのは難しく、希望する特徴を記述に反映することはできていない。

値動きは多峰の山/上昇/下降の階層的組合せからなる解析木で記述される。この記述方法を過去のデータに適用し、共通する解析木のパターンを発見し、予測に用いることを検討した。解析中のフロンティアを値動きの記述として、特徴量とし次時点の値動きの予測を試みた。しかし、統計的にはほぼランダムであり、予測精度の向上は図れなかった。

次に、従来方法の改善と同時に、経済時系列予測時における、当該価格データ以外のデータの有効性、非線形表現手法の有効性に關

する検討結果を記す。

一般に経済時系列として考察する、外国為替交換レート、株価等の金融商品の市場は、極めて効率的と考えられ、時折入る外部からの情報に対する瞬時的な応答を除けば、自律的(独立増分を持つ確率過程)に動いているように見える。しかし、実際には、瞬時的な応答は、瞬時的ではなく当該情報に対する影響は暫時残り、増分は独立ではなく、僅少ではあるが、「以前の動作・状態」に依存している。

本報告前半の記述は、この「以前の動作・状態」をトレンドという形で記述しようという試みである。後半の記述は、この「以前の動作・状態」は、人間の見た目のパターンとして記述するのは難しく、単純なしかし非線形な記述によって可能となるという仮説の検討の結果である。

考慮した特徴量は、予測期間、すなわち過去データをサンプルする期間を複数個として得られる過去データ、他の類似の金融商品価格、ブログ投稿であり、考慮した手法としては、異なる特徴量には異なる特徴関数を割当てた上で予測関数を記述する MKL (multiple kernel learning)、離散関数を予測関数とできる遺伝的アルゴリズム、離散関数と連続関数の混在を許す予測関数を用いる差分進化計算(differential evolution)等である。

対象とする価格としては、外国為替変換レート、株価、原油価格を試みた。

なお、評価方法としては、通常よく使われる値動き方向の予測と実際値の当り数割合である hit ratio、平均二乗誤差以外に、実際に近い取引戦略(取引戦略自体も学習対象である)を用いて得られる利益、無リスク資産と volatility を考慮した Sharpe ratio 等を用いた。評価基準を金融取引に似せた理由には、価格の値動き自体が利益やヘッジを目的とする金融取引によって惹起されたものであること、手法・特徴量による利益の順位差と hit ratio や平均二乗誤差による順位差とが必ずしも関連しないからである。

以下では、これらの指標を予測精度と呼びその定性的な記述に留める。

他市場の影響がある、言い換えれば、他市場の値動きを参照すると予測精度が向上することを確認した。

外国為替交換レートの場合、例えば、米ドル円レートの予測に際して、英ポンドドルや欧ユーロドルのレートの値動きを特徴量とすると予測精度が向上する。ただし、これらの値は相互に関連しているため、原因・結果の方向は両方あり、時刻・状況によって異なる。そのため、予測関数は線形ではありえない。これらが見かけの影響でないことを確認するため、説明変数として導入した他の為替レートを類似の統計的性質をもつランダム系列におきかえて予測をさせると予測精度が悪化することを確認している。ただし、ど

のように影響しているかの記述は、困難であった。複数個の実験、また同一の実験であっても時刻を変えて、得られた予測関数の構造・パラメータを調べたが、ばらつきが大きく、特に、遺伝的アルゴリズム・進化計算を用いた場合には全く知見が得られなかった。

原油価格の場合、WTI 価格と Brent 価格について実験を行った。外国為替交換レートと同様に、他市場の価格の動きを説明変数として用いることにより予測精度が向上した。ただし、外国為替交換レート同様に、因果関係は固定・単純ではなく、状況によって変化していると考えられる。また、ランダム系列によって代替し、ランダム系列では予測精度が向上しないことを確認している。

ブログを用いて株価の動きの予測精度を向上させることができた。この研究の興味深い結果の一つは、予測精度の改善がブログの内容には依存していないことである。理由としては、ブログ内容の解析が難しく、従って誤りやすいために、それを説明変数とすることに意味がなかったと考えられる。一方、実際に効果があると検証できたのは、ブログの件数である。これが効果を示したのは、推測であるが、値動きに遅延が存在するからである。すなわち、件数が多ければ、遅延した値動きが継続・拡大し、一方、件数が少なければ、遅延した値動きが早期に終了するからである。言い換えれば、ブログの内容が遅延した値動きの初期の動きに現れるからである。

時系列予測においては、異なる予測時間 (time horizon/frame) の予測を組み合わせることに効果があると推測できる。すなわち、例えば、一時間後の値動きは、より短い時間、30 分や 10 分の市場の動き、より長い期間、2 時間、6 時間、一日の市場の動きに影響されていると推測することは容易にできる。

実際、今回、外国為替交換レート、原油価格の予測に用いたが、弱くはあるが明確な影響があった。すなわち、予測したい time horizon とは異なる time frame のデータが予測精度の向上に祐子であった。

予測関数の表現に用いた MKL は、SVM/SVR において、kernel 関数として、複数の kernel 関数の荷重付き和を用い、その荷重も合わせて学習する枠組みである。非線形さは特徴量によって異なるため、また被説明変数への影響度合いは特徴量によって異なるため、このような枠組みで説明されうる現象は多いと考えられる。

実際、今回は、外国為替交換レート、株価、原油価格において MKL を用いたがいずれも、有効である、すなわち、予測精度の向上に貢献していることが確認できた。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 4 件)

(査読有) Deng, S., Yoshiyama, K., Mitsubuchi, T. & Sakurai, A. Hybrid Method of Multiple Kernel Learning and Genetic Algorithm for Forecasting Short-Term Foreign Exchange Rates. Computational Economics. Vol. 45, No. 1, pp.49-89 (2015).

(査読有) Deng, S., Mitsubuchi, T. & Sakurai, A. Stock price change rate prediction by utilizing social network activities. The Scientific World Journal, Vol. 2014, 861641 (2014).

(査読有) Deng, S. & Sakurai, A. Integrated model of multiple kernel learning and differential evolution for EUR/USD trading. The Scientific World Journal, Vol. 2014, 914641 (2014).

(査読有) Deng, S. & Sakurai, A. Crude oil spot price forecasting based on multiple crude oil markets and timeframes. Energies, Vol. 7, No. 5, pp. 2761-2779 (2014).

[学会発表](計 1 件)

(査読有) Deng, S. & Sakurai, A. Foreign exchange trading rules using a single technical indicator from multiple timeframes. Proceedings of 27th International Conference on Advanced Information Networking and Applications Workshops, WAINA 2013, Barcelona (Spain), pp. 207-212 (2013).

6. 研究組織

(1) 研究代表者

櫻井 彰人 (SAKURAI AKITO)
慶應義塾大学・理工学部・教授
研究者番号：00303339

(2) 研究分担者

該当なし

(3) 連携研究者

該当なし