

時系列パターン抽出に基づく個人投資家意思決定支援システムの実現

An Individual Trader Decision Support System with Time-Series Patterns

杉山 喜昭*¹
Yoshiaki Sugiyama

平林 悟*¹
Satoru Hirabayashi

阿部秀尚*²
Hidenao Abe

山口高平*¹
Takahira Yamaguchi

*¹ 慶應義塾大学
Keio University

*² 島根大学
Shimane University

This research intends the implementation of a system to discover a novel rule based on technical indices by the learning of the time-series data of the technical indices values. The overview of the system is as follows: Sub-sequences are extracted from the time-series date of the technical indices values first, and then a clustering algorithm is used on the Sub-sequences to find representative patterns, and lastly if-then rules consisted of the representative patterns are outputted as trading-rules by using decision tree learning. At the same time, in order to validate the trading-rules, the verified methodology is researched in this paper. By analyzing the learned rules, a trader can acquire novel knowledge, which can lead to enable him to broaden his viewpoints.

1. はじめに

本研究では、株価を基礎として、テクニカル指標値の時系列データから有用な時系列ルールを発見し、そのルールを検証する環境を提供することで、個人投資家の意思決定を支援するためのシステムの実現を目的とする。先行研究[1]ではテクニカル指標値の組み合わせによって売買ルールを作成したのに対し、本研究においてはテクニカル指標値の時系列パターンを組み合わせることによって売買ルールの作成を行う。

提案アプローチは、時系列ルールの発見では、テクニカル指標値の時系列データからクラスタリングアルゴリズム及び決定木学習を用いて売買ルールを出力する。検証環境では、過去の株価データの中から売買日と株価推移が類似した期間を複数算出し、それらの期間において売買ルールを適用させた場合の売買評価を出力する。

2. 時系列データにおける特徴抽出

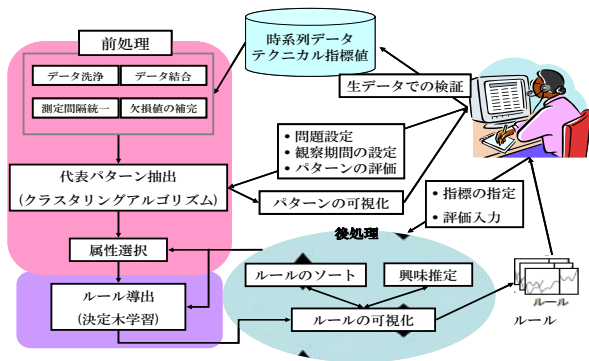


図1 システム概要

時系列ルールの発見手順は、テクニカル指標値の時系列データから複数のパターンを抽出し、それらのパターンについてクラスタリングアルゴリズムを用いることによって代表的なパターンを見つけ出し、それらの代表パターンを決定木学習により売買ルールとして出力する。

2.1 代表パターン抽出

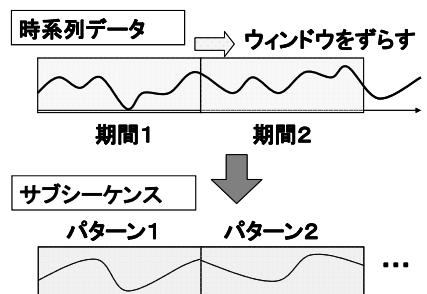


図2 パターンの切り出し

上図2のように、テクニカル指標値の時系列データからウィンドウサイズに応じたサブシーケンスを切り出す。ウィンドウサイズはサブシーケンスを抽出するための観察期間として何日か、ということの意味する。このようにして、サブシーケンスを各テクニカル指標値において切り出す。テクニカル指標名を属性としてCSVで表現したデータベースからテクニカル指標(TI)の値を縦に見ていったものがその期間に抽出されたパターンとなる。(図3)

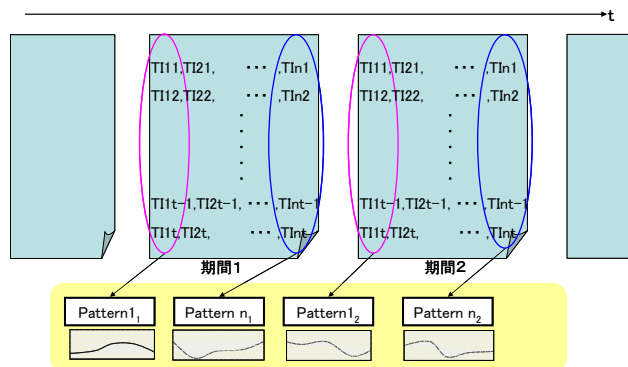


図3 時系列データからのパターン抽出

そして、切り出されたパターンに対してクラスタリングを行い、任意の数の代表パターンを抽出する。

2.2 ルールの導出

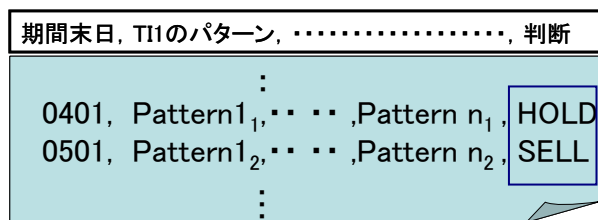


図4 訓練データの形式

図4のように、各期間に対して、テクニカル指標ごとに抽出された代表パターンを属性とし、期間末日の株式売買における正しい判断をクラスとして、その集合を訓練データとする。ここで、正しい判断とは設定された百分率(x)に対して、その日から20日間で終値がx%増加したら「買い」、x%減少したら「売り」、そのいずれでもなかったら「保持」を選択することをいう。ただし、「買い」と「売り」を出すための条件とをともに満たした場合は、増減率の大きいほうを採用することとする。

作成されたデータについて、決定木学習アルゴリズムを用いた場合、図5のようなルールを得ることになる。時系列ルールを訓練データとは異なるテストデータに適用することで評価を行うことが可能となる。

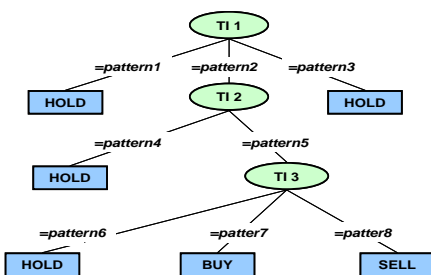


図5 導出されるルール

2.3 ケーススタディ

本システムについて、実際の株価データを使用してシミュレーションを行い評価する。

(1) 実験方法

株価データはカブロボから取得した[2]。カブロボは株式自動売買ソフトウェアである。取得期間は以下のとおりである。

- 訓練データは2004年2月2日～2005年12月30日
- テストデータは2006年1月4日～2006年12月29日

テクニカル指標は17種類用いた。各テクニカル指標値の算出に必要な日数などのパラメータの値は一般的に用いられることの多い期間を設定した。当然個人投資家のスタンスにより、パラメータは随時変更可能である。

代表パターンを抽出するための観察期間は10日(最大8営業日)、パターン数は4から最大で10とした。

他の主な条件は以下のとおりである。

- 資金:5000万円
- 1日1回の取引を行う
- 一度に注文する株数は単位株数とする

今回は業種単位でルールを作成した。用いた業種は表1のとおりである。

表1 使用銘柄

業種	銘柄	単位株数
金融	三井住友フィナンシャルグループ	1
	みずほフィナンシャルグループ	1
通信	KDDI	1
	NTTドコモ	1
	ソフトバンク	100
鉄道	東日本旅客鉄道	1
	東海旅客鉄道	1

(2) 実験結果

シミュレーションの結果に基づき、業種ごとに損益を出す。また、時系列ルールによる結果を比較するためランダムで売買を行った場合の損益を出す。ここでは損失の大きかった三井住友フィナンシャルグループと、利益の多かったKDDIとソフトバンクの結果を図6、図7に載せる。

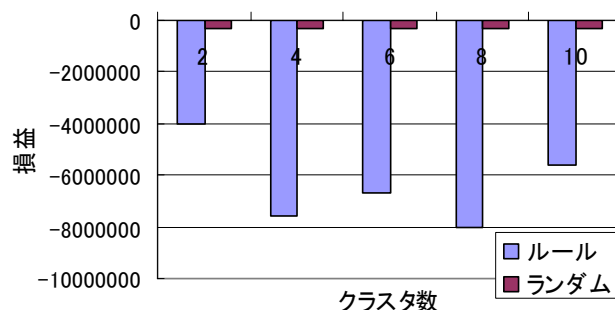


図6 三井住友フィナンシャルグループ

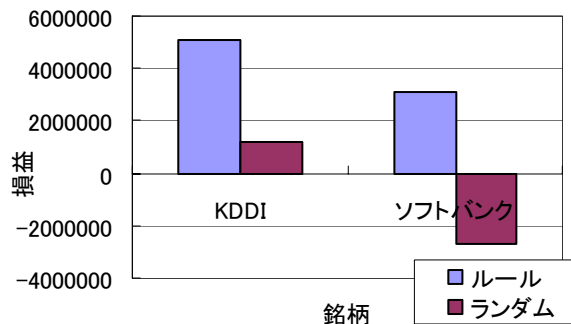


図7 KDDIおよびソフトバンク

図7におけるクラスタ数はKDDIが4、ソフトバンクが8である。

(3) 考察

図6の結果から大きな損失の原因について考察する前に、損益と期間内最大資産および最小資産についてグラフ化したものを図8に載せる。

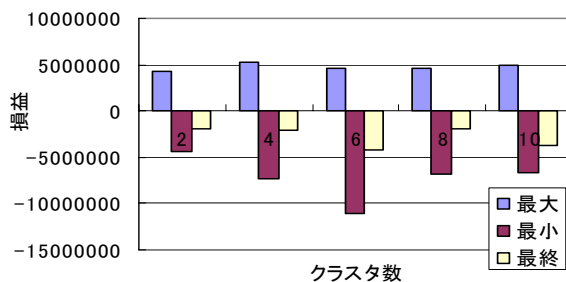


図8 三井住友フィナンシャルグループ最大・最小利益

図 8 から最大資産の値が、クラスタ数を問わず 8%以上の利益を一時的に上げていて、ある程度高いことが分かる。つまり、クラスタごとに作成されたルールセット全体で見るとはならず、ルールセット中の単独ルールの中には、有益な働きをしているものが存在することを示唆していると考えられる。しかしながら、今回は明らかにマイナス方向に進んでいる度合いが大きい。その原因としてはテスト期間に用いられている企業の株価の変動を考慮に入ると、現状にそぐわないルールによって売買される頻度が高いことが考えられる。つまり、不要なルールを取り除くことが必要である。そのためには、枝狩りを行うなどの後処理が必要になってくると考えられる。

図 7 では、両銘柄共にルールセットがうまく機能していると考えられる。とりわけ KDDI では売買後利益が 500 万円となり、初期資産 5000 万円から考えると利益は 10%である。ルールセットをさらに改善することで先ほどの三井住友フィナンシャルグループより積極的な意味でよりよいシミュレーション結果が得られると考えられる。

次に作成されたルールについて考察を行う。KDDI のクラスタ数 4 のときのルールセットの中で 1 つ目について考える。ルールは以下のようにになっている。

Volume=1 かつ volumeratio=3 かつ bollingerbandup=1 ならば SELL ...①

ここで上記の番号はその指標のパターン番号を示す。それぞれの指標のパターンを以下に示す。

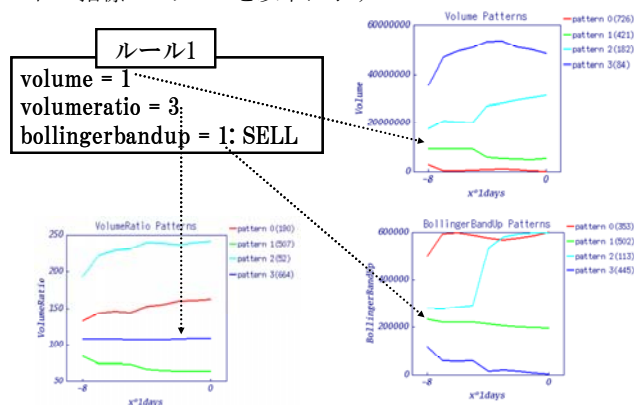


図 9 各テクニカル指標のパターン

図 9 から上記ルール①は、Volume(取引量), volumeRatio, BollingerBandUp が小さくなっているときには売らなければならないことを示唆している。これを解釈すると、volumeRatio は取引量と相関があるが、取引量が減り、トレンドを示す BollingerBandUp が下降傾向にあるという状況、要約すると、この銘柄(KDDI)の人气が落ちているときには、本銘柄を売るべきであるということであり、個人投資家にとっても直感として納得できるものである。今回は投資家の直感にそぐうルールとなっているが、投資家の直感にはない新しい発見や知識が見つかる可能性を秘めていることが分かる。

3. 検証環境支援

3.1 検証環境システム

個人投資家が、機械学習によって提示されたルールを実際に使用する前に、過去のデータでルールの有用性を確認したいと考えることは必然である。そこで、新規に発見されたルールを容易に検証することが可能である環境を提供することが必要

となる。個人投資家が検証を行う第一の目的は、売買を行う日(当該日)の株価水準で自分の売買意思決定はどのくらいの信頼性があるのかについて知ることである。

本章で提案する検証環境では、現在の株価水準と類似した過去の株価水準を探し当て、その期間において組み合わせルール[1]を適用し、ルールの評価を行う。ただし、組み合わせルールとは 2 章の出力である時系列ルールではなく、過去のルール作成の研究によって出力される組み合わせルールを採用した[1]。過去の株価データを用いて、ある日の各テクニカル指標値とその日にすべきであった売買判断を 1 日ずつずらして記録し、決定木学習によって If Then ルールを作成するものである。2 章のシステム作りと平行開発のため過去の研究を用いるが、今後は段階ごとに一貫したシステムを再構築していく。

下図 10 が本検証環境システムの全体図である。

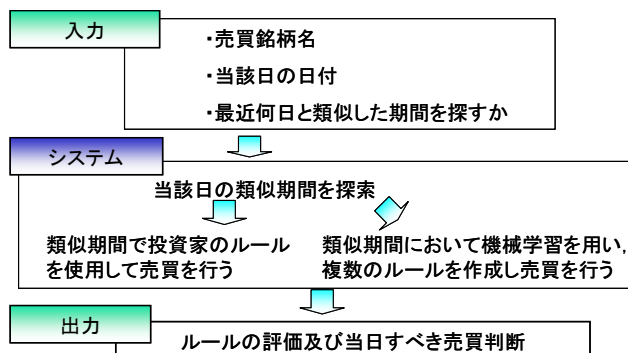


図 10 検証環境概要

株価水準が類似した過去の期間を算出する具体的な方法は、当該日前数十日間(類似探索期間)の株価と過去の株価の差を二乗したものの総和が小さいものから類似期間として複数求める。最も類似した期間近辺において組み合わせルールを作成し、他の複数の類似期間においてルールを用いて売買シミュレーションを行い、評価指標の値を算出するとともに、いつどのような売買を行ったかチャート図(図 11)による出力を行う。類似期間においてルールを作成する意義は、当該日と類似した期間で学習させることにより、適当な期間で作成するルールに比べ現状に合ったルールを作成できると考えるからである。また出力結果が個人投資家の投資スタンスに合わなかった場合を考慮し、売買銘柄による単一の組み合わせルールだけでなく、売買銘柄と同業種である他銘柄による複数の組み合わせルールを作成し評価を行うことで、個人投資家の選択の幅を広げることが可能となる。同業種内でルールを作る根拠として、株価の平均値が類似していること及び各業種は社会的要因によって同様の株価の変化を示す可能性が高いことが挙げられる。[4]

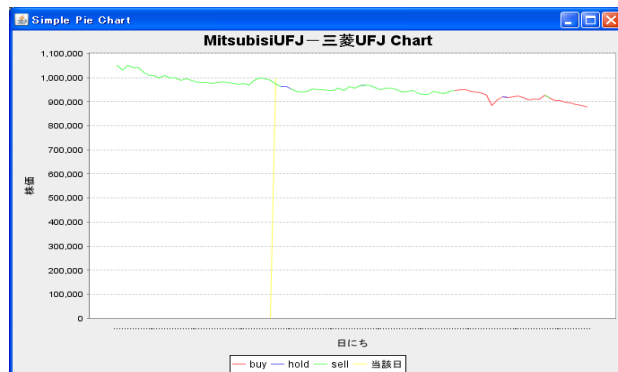


図 11 売買判断チャート表示画面

図 11 では緑色の日が「売り」、赤色の日が「買い」、青色の日が「保持(何もしない)」という判断をルールが行ったことを示す。個人投資家は、複数作成された組み合わせルールを適用した売買シミュレーション結果からチャート及び、評価指標の値を基に最も自分の投資スタンスに合ったルールを容易に判断することが可能となり、意思決定の支援となる。

3.2 ケーススタディ

本システムについて、実際の株価データを使用してシミュレーションを行い評価する。

(1) 実験方法

株価データは 2.3 と同様カブロボから同期間分取得した。株式売買を行う現在日を 2006 年 10 月 4 日とする。また、類似探索期間を 30 日とする。最も類似した期間において組み合わせルールを作成し、2, 3 番目に類似した期間および現在においてルールを適用する。テクニカル指標は 2.3 と同様であり 17 種類用いた。

他の主な条件は以下のとおりである。

- 1日1回の取引を行う
- 一度に注文する株数は単位株数とする
- 信用売りも可能

今回は銘柄単位でルールを作成した。用いた業種は以下のとおりである。

表 2 使用銘柄

業種	銘柄	単位株数
金融	三菱UFJフィナンシャルグループ	1
	三井住友フィナンシャルグループ	1
	みずほフィナンシャルグループ	1
通信	KDDI	1
	NTTドコモ	1
	ソフトバンク	100
	NTT	1
鉄道	東日本旅客鉄道	1
	東海旅客鉄道	1
	商船三井	1000

(2) 実験結果

シミュレーションの結果に基づき、銘柄ごとに損益を出す。ここでは金融業の結果を図 12 に載せる。

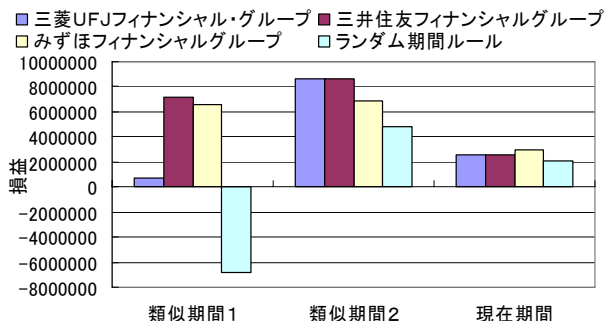


図 12 金融業

図 12 は、三菱UFJフィナンシャルグループにおいて現在期間と最も類似した過去の期間に対し、金融業各銘柄が各々の株価データを用いてルールを作成し、他の2つの類似期間お

び現在期間の三菱UFJフィナンシャルグループの株価データに適用した場合の損益結果である。

(3) 考察

図 12 の結果から、多くの個人投資家はそれぞれの類似期間で良い成績を残した三井フィナンシャルグループのルールを現在期間に使用することが想定される。実際、現在期間においてもこのルールは利益を上げていることが確認できた。一方、ランダム期間において作成したルールでは不安定な成績であり、利益を上げて他に比べ少ないことがわかる。よって、個人投資家がこのランダム期間ルールを信用して使用することは難しいと考えられる。類似期間においてルールを作成する有用性が確認できた。

三井住友フィナンシャルグループの中で最もよく使われ成績もよいルールとして、以下のルールが抽出できた。

HLBandDown > 680000 かつ BollingerBandUp > 752470 ならば SELL

上記ルールはどの期間においても正解率が高いため、今後期待できるルールであることが分かった。

このように個人投資家は、与えられたルールをただ現在期間に適用するのではなく、同業種で作成された複数のルールの中から、過去の類似期間でのシミュレーションによる損益や評価指標値を基に、自分の投資スタンスと合うものを選択することができると考える。

4. おわりに

本研究では、テクニカル指標値の時系列データを用いたパターン抽出に基づく個人投資家意思決定システムの提案を行った。数種類の業種の株価データを用いたケーススタディを通して、時系列パターン抽出では初期資産に対して 10% の利益を得た銘柄も存在し、早計かもしれないが一応の評価を示した。検証環境支援においては類似期間にルールを作成し、類似期間においてルールの検証を行うことが有用であることを示した。これにより、機械学習によって得られた時系列ルールを投資家に信用してもらう材料を与えることができた。総じて、本研究により、個人投資家の労力を軽減すると共に、新たな知見を得る可能性を秘めており、それによって、株式市場についてより理解し、深い考察ができるようになってくる。しかしながら、ケーススタディを通して、問題点も浮き彫りになってきた。業種固有のルールをその業種に属する銘柄に適用したときに大きな損失がでるといことはそのルールの作成に問題があるということである。また、抽出されたパターンを見て、生データとの乖離が大きかったことも問題と考える。今後の展望としては、データマイニングの過程で発見された知識であるルールの後処理や、機械学習に用いる手法の再検討を行う。また、検証環境を発展させ時系列パターン抽出によって得られた時系列ルールも検証できるようにし、システムの一貫性を目指す必要がある。

参考文献

- [1]内藤憲吾, 平林悟, 阿部秀尚, 山口高平, テクニカル指標組み合わせルールの学習に基づく投資家意思決定支援, 第 21 回人工知能学会全国大会, 3B8-4, 2007
- [2]海鳥不二夫, 自動株式売買ソフトウェア・スーパー株ロボを作ろう! 秀和システム, 2006
- [3]元田 浩, 山口 高平, 津本 周作, 沼尾 正行, データマイニングの基礎, オーム社, 2006
- [4]林康史, 基礎から学ぶデイトレード, 日経 BP 社, 2005