

## 리프집합을 활용한 캔들스틱 트레이딩 최적화 전략

김현호<sup>1</sup> · 오경주<sup>2</sup>

<sup>12</sup>연세대학교 정보산업공학과

접수 2012년 5월 31일, 수정 2012년 6월 25일, 게재확정 2012년 7월 27일

### 요약

본 논문에서는 선물시장에서 리프집합과 의사결정나무를 이용한 매매규칙 기반의 시스템 트레이딩 전략을 제안한다. 과거 데이터마이닝 방법론을 이용한 선물시장 투자전략에 대한 많은 연구가 진행되어 왔으나 상대적으로 다양한 변수의 조합을 통한 시스템 트레이딩에 대한 연구는 거의 없었다. 본 연구는 크게 세 가지 목적을 가지고 있다. 첫 번째 목적은 매매규칙 기반 시스템 트레이딩에서 의사결정나무 방법론의 사용이 투자성과에 어떠한 영향을 미치는가를 분석하는 것이다. 두 번째 목적은 단기매매부터 장기 매매까지 중에서 적절한 매매 시간간격을 찾아내는 것이다. 세 번째 목적은 매매규칙 생성 시 사용되는 최적의 트레이닝 구간을 찾는 것이다. 이 논문의 실험결과는 제안한 투자전략의 유용성을 증명할 수 있을 것이며, 또한 이를 통해 시장참여자들에게 투자 결정에 있어 도움을 줄 수 있을 것이다.

주요용어: 리프집합, 선물시장, 시스템 트레이딩, 의사결정나무, 최적화 트레이딩.

### 1. 서론

주식시장에 참여하는 투자자들에게 있어서 주가의 향방에 대한 예측은 투자에 의해 얻을 수 있는 수익을 결정한다. 그에 따라 시장에 참여하는 모든 투자자들은 주가의 향방을 예측할 수 있는 방법을 강구하게 되고 각종 이론과 시장정보, 그리고 투자자 자신의 경험에 비추어 다양한 예측기법을 적용하려 한다. 그러한 예측 방법론으로 초기의 주가지수에 대한 연구는 주로 시계열분석을 통한 통계적 방법을 이용하여 주가지수를 분석하여 왔고, 판별분석이나 로지스틱 회귀분석 등 여러 가지 통계적 분석 방법을 이용하여 예측하려는 시도가 계속되어 왔다. 정보의 누적으로 인하여 데이터가 광범위해졌고, 이에 따라 데이터마이닝이라는 큰 범주 안에서 인공지능망이나 사례기반추론, 유전자 알고리즘 같은 방법을 이용하는 방법론이 연구되어 왔다.

다만 이 연구에서는 주식이 아닌 선물시장을 대상으로 한다. 선물시장의 투자를 위한 좋은 조건을 살펴보면, 첫 번째로 거래 비용적 측면을 보면 우선 수수료의 관점에서 보통 거래건수에 해당되는 일 정금액으로 현물 수수료보다 훨씬 적은 비용으로 투자할 수 있는 여건이라고 할 수 있다. 그리고 사고 팔 때의 가격차이가 없다는 점, 마지막으로 가격효과가 현물의 5%밖에 되지 않는다. 또한 레버리지 효과도 현물보다는 선물에서 좀 더 큰 레버리지 효과를 볼 수 있으며 실제가격의 15%의 증거금으로 거래를 할 수 있다는 점에서 초기자본의 크기가 다른 것이 그 원인이 된다. 이러한 몇 가지 이유로 투기적인 목적으로 선물 시장에 투자하는 것이 현물 시장에 투자하는 것보다는 의미가 있다.

<sup>1</sup> (120-749) 서울특별시 서대문구 신촌동 134번지, 연세대학교 정보산업공학과, 석사과정.

<sup>2</sup> 교신저자: (120-749) 서울특별시 서대문구 신촌동 134번지, 연세대학교 정보산업공학과, 부교수.

E-mail: johanoh@yonsei.ac.kr

이런 방법론을 사용한 기존의 연구들은 끊임없이 변화하는 시장상황에 적용하기 어렵다는 한계를 갖는다. 선물시장을 비롯한 금융시장에서는 과거에 발생한 추세나 패턴이 미래에도 반복된다고 단정할 수 없다. 따라서 일관된 특정한 투자전략이 꾸준히 같은 성과를 보장하지 않는다. 과거 데이터에 근거하여 놀랄만한 수익을 보여주는 투자전략도 실제로 시장에 적용시키면 그 비효율성이 여실히 드러난다. 과거 데이터는 이미 확정된 데이터이므로 높은 수익을 보장하는 적합한 전략을 구하기가 쉽다. 과거 데이터에 대한 과적합 (overfitting) 현상 때문이다. 그러나 미래의 데이터는 과거와 같은 패턴으로 확정되지 않고, 전혀 예측이 불가능하다는 것이 정설이므로 과거에 사용되었던 특정한 전략에 대해 앞으로도 같은 수익률을 기대할 수는 없다. 즉, 기존의 연구들은 과거 데이터에 적합화된 특정한 투자전략을 사용하기 때문에 시시각각 변화하는 시장상황에 알맞은 퍼포먼스를 보일 수 없다는 한계점을 갖는다. 따라서 본 논문에서는 시장의 다양한 패턴과 추세가 포함되는 데이터를 사용하여 리프 집합을 이용한 다양한 조합의 투자전략을 생성한다. 상승, 하락, 횡보의 다양한 시장 패턴이 내포된 데이터 구간의 시뮬레이션을 통해 시장에서 발생하는 다양한 패턴과 추세에 대한 투자전략의 성과를 분석한다. 이를 통해서 특정한 과거 데이터가 아닌, 다양한 시장 데이터에 의해 검증된 투자전략을 찾는 것이 이 논문의 중점이다.

본 논문은 한국 주가지수 선물 (KOSPI200)의 가격 데이터를 기반으로 의사결정나무와 리프집합을 이용하여 매매규칙을 찾아내어 이 규칙을 적용시켜 거래를 하였을 경우에 투자자가 어느 정도의 수익을 낼 수 있는가를 측정하는 것이 목적이다. 나아가 투자성과가 높은 투자 전략을 찾아내어 최적의 매매규칙을 찾는 것이 본 논문의 목적이다.

현실적인 분석을 위하여 백테스팅은 실제적인 슬리피지와 거래비용을 반영한다. 슬리피지는 시장에서의 형성되는 매수, 매도 호가와 실제적으로 거래가 이루어지는 가격간의 차이를 뜻한다. 슬리피지는 유동성 공급자, 주문 수량, 현재 호가의 누적수량 등의 다양한 호가 형성 요인들에 의해서 결정되며 대부분의 주문에서 발생하는 손실이다. 거래비용은 거래수수료, 중개수수료, 직접세 등이 포함되는 거래 시 발생하는 비용이다. 매 거래 시 슬리피지와 거래비용은 투자자의 수익을 감소시킨다. 그러므로 빈도 높은 투자전략은 그렇지 않은 투자전략에 비해 수익 감소의 폭이 더 크다. 따라서 슬리피지와 거래비용의 반영은 투자전략의 성과분석에 큰 영향을 미친다.

본 논문은 다음과 같이 구성되어 있다. 2절에서는 기존의 투자전략에 대한 선행 연구를 분석 하였으며, 3절에서는 이 논문에서 제시하는 투자전략에 대하여 자세히 설명하고, 4절에서는 제안한 투자전략의 실증분석 결과를 분석하였으며, 마지막으로 결론에서는 본 연구의 기대효과 및 향후 연구에 대해 서술하였다.

## 2. 연구배경

### 2.1. 캔들차트 전략

18세기 Munehisa Homma에 의해서 제안되고 Nison (1991)에 의해 널리 알려진 캔들차트는 가격의 움직임을 촛대모양으로 표기하여 시각화 하는 기법이다. 캔들차트는 주어진 시간 간격 동안의 가격 움직임을 선과 막대로 표기하므로, 라인차트와 바차트의 결합된 형태로 볼 수 있다. 하나의 캔들은 주어진 시간 간격 동안, 거래가 체결되는 가격에 대한 시가, 고가, 저가, 종가의 정보를 담으며, 시가와 종가로 생성되는 가격 움직임의 방향까지 나타낸다. 캔들스틱은 가격의 움직임을 직관적인 시각적 표현으로 제공하므로 다양한 분야에서 사용되고 있으며, 기술적 분석을 위한 전제 조건이다.

## 2.2. 기술적 분석

기술적 분석은 주식시장을 비롯한 금융시장을 분석하고 예측하는 기법으로 가격 데이터를 사용하는 분석기법이다. Cutler (1989)가 정의한 경제, 산업, 기업환경을 기반으로 하여 주가의 내재적 가치를 결정하는 기본적 분석과는 반대로, Murphy (1999)에 의하여 정리된 기술적 분석은 그래프로 나타내어지는 가격, 거래량의 역사적 데이터를 사용하여 미래 가격 움직임을 예측하기 위해 사용되는 도구이다.

기술적 분석은 주가를 평가하기 위한 두가지 기법을 사용한다. 첫 번째 기법은 시가, 고가, 저가, 종가와 거래량으로 이루어지는 기초적인 그래프를 이용하는 방법이다. 다른 방법은 매매결정을 내리기 위한 계산이나 수학적 과정을 거치는 기술적 지표를 사용하는 방법이다. 주가는 기본적으로 주식의 내재가치에 의해 결정된다. 반면에 주식 선물 시장은 주가 움직임의 예측에 대한 리스크 헷지를 목적으로 형성되었기 때문에, 주식 가격 변동을 정확하게 예측하는 것이 선물시장에서의 거래를 위해 매우 중요하다 (Lee, 2010). 그러므로 주식 선물 시장에서 예측에 대한 기술적 분석은 내재 가치 평가보다 큰 주목을 받는다. 기술적 분석에 사용되는 지표는 1800년대 Brock (1992)의 기술적 지표부터 시작하여 다양한 기술적 지표가 현존한다. 기술적 지표는 투자에 대한 시기와 카테고리를 결정하는데 도움을 주기 위해 사용되는 변수들이며, 시가, 고가, 저가, 종가, 거래량을 고려하여 계산된다.

## 2.3. 러프집합

러프 집합 이론은 1982년에 Pawlak에 의해 처음 제시되었다 (Pawlak, 1982). 이 이론에 쓰일 수 있는 부분은 속성들 간의 상관성을 찾는 부분과 어떤 속성의 중요성을 파악하는 것, 그리고 일관성이 없는 데이터를 처리하는데 사용하는 것들이다. 이 접근 방법은 불확실성으로 가득한 데이터를 처리할 수 있다는 점에서 확률이론, 증거이론, 그리고 퍼지이론과 공통점을 가진다.

이 이론은 세상에 모든 개체들은 그들이 가진 어떤 정보로써 집합을 지을 수 있다는 가정 하에서 시작된다. 동일한 정보의 외연을 가진 개체들은 그들의 정보로 인해 동일한 것으로 취급된다. 이러한 동질성 관계 (indiscernibility relationship)가 러프 집합 이론의 기초가 된다.

어떤 동질성을 가진 집합을 기본집합 (elementary set)이라 하고 이들이 어떤 전체집합에 대한 지식의 기본 단위를 이룬다. 어떤 기본집합의 합집합이 되는 집합을 일반집합 (crisp set)이라 하고 그렇지 않은 경우를 러프 집합 (rough set)이라 한다. 이로 어떤 집합은 그 집합이나 여집합의 구성원으로 확실하게는 속하지 않을 수 있는 개체들 즉, 경계부분개체 (boundary-line cases)를 가진다.

이로 러프 집합은 하위근사와 상위근사로 불리는 몇몇의 일반집합으로 표현될 수 있다. 전자는 확실성 있게 한 집합에 속하는 경우이고, 후자는 목표로 하는 집합에 속할 수 있는 경우라 할 수 있다.

## 2.4. 의사결정나무

의사 결정 나무는 분류 (classification)을 수행하는 통계학적 모델의 하나로 의사결정 규칙을 나무구조로 도표화하여 분류와 예측을 수행하는 분석방법이다. 해석이 직관적이고 용이하며, 범주형 변수 (categorical variable)를 다루기 쉽고, 설명 변수 (explanatory variable)의 차원이 높아 관찰 값의 개수보다 관찰 값 항목의 숫자(즉, 설명변수의 개수)가 많아도 분석이 가능하다는 점에서 실생활에서 활용도가 높은 분석 모델이다. 나무를 생성하는 알고리즘에 따라 CART (Breiman 등, 1984), C4.5 (Quinlan 1993), QUEST (Loh와 Shih, 1997), CRUISE (Kim와 Loh, 2001) 등의 모델이 제시되었다.

의사 결정 나무의 생성은 (1) 분리 변수 선택 (variable selection)과 분리점 선택 (split point selection), (2) 분류 및 예측, (3) 가지치기 (pruning)를 통한 나무의 크기 결정의 단계로 진행된다. 각

단계를 해결해 가는 방식의 각기 다른 조합의 알고리즘이 각기 다른 방식의 의사 결정 나무 모델로 제시된다.

### 3. 제안한 투자전략

본 절에서는 제안한 투자전략의 구조와 특징에 대해 설명한다. 제안한 투자전략의 1단계에서는 선물 체결 틱 데이터를 다양한 시간 간격의 기초 데이터로 변환한다. 2단계에서는 러프집합과 의사결정나무를 이용하여 매매규칙을 정규화하여 트레이딩 규칙 (rule)을 생성한다. 그리고 마지막 단계에서는 선물시장에서의 백테스팅을 통한 매매규칙의 성과를 분석한다.

#### 1단계. 선물 원시 데이터를 다양한 시간 간격의 기초 데이터로 변환

본 논문에서는 다양한 시간 간격의 선물 트레이딩에 대한 성과를 분석하여 최적의 매매 간격을 분석한다. 이를 위해 동일한 기간의 선물 데이터를 다양한 시간간격의 데이터로 변환한다. 선물시장의 가장 기본적인 데이터인 원시 데이터를 여러 가지 다양한 시간간격의 데이터로 변환한다. 다만 시간간격이 짧으면 매매 빈도가 높아져 거래비용으로 인한 손실이 커지게 되며, 시간 간격이 길면 매매 빈도가 낮아져 의미 있는 매매 기회를 포착하지 못하게 되므로, 의미가 있는 시간 간격에 대한 데이터만을 사용한다.

#### 2단계. 러프집합과 의사결정나무를 이용하여 매매규칙을 정규화

전 단계에서 생성된 각각의 시간 간격 데이터에 대해 독립적으로 매매규칙을 생성한다. 매매규칙을 생성하는 과정은 데이터마이닝 방법론의 트레이닝 과정이므로 과거 선물시장의 일정 부분을 선택하여 적용한다. 즉, 과거 선물시장의 일정한 트레이닝 기간에 대한 기술적 지표 값을 계산한다. 기술적 지표 계산 값의 범위는  $(-\infty, \infty)$ 의 광범위한 범주를 갖지만 의미 있는 구간은 한정되어 있으므로, 유효한 범위 내에서 이산화 하여 구간을 구분한다. 이산화 구간을 선택하여 매매규칙을 생성하는 방법은 일반적인 트레이딩 시뮬레이션 방법과 의사결정나무의 두 가지 방법을 사용한다. 본 논문의 목표인 의사결정나무의 성과 개선 효과를 분석하기 위해, 각기 방법론을 적용하여 매매규칙을 생성한다. 매매규칙을 생성하기 위해 다양한 기술적 지표를 사용하게 되는데, 기술적 지표를 중첩하여 적용하게 되면, 기술적 지표의 개수에 따라 비례하여 매수, 매도 신호의 조건이 엄격해지므로 신호가 발생하기 어렵게 된다. 따라서 전체 기술적 지표들 가운데 수익을 내는데 알맞은 효율적인 소수의 기술적 지표만을 선별해야 한다. 이 기술적 지표의 선별과정에 일반적인 트레이딩 시뮬레이션 방법과 의사결정나무의 방법론을 적용하여 각 방법론의 성과분석을 실시한다.

일반적인 트레이딩 시뮬레이션 방법을 적용한 매매규칙의 생성은 다음의 과정을 따른다. 다양한 기술적 지표에 대한 계산을 통해 발생하는 매매신호에 따라 트레이닝 구간에서의 수익률을 비교하여 상위의 수익률을 내는 기술적 지표를 선택한다. 이 선별된 기술적 지표에 대하여 이산화 구간을 정한다. 여기서는 이산화 구간을 정하는 기준으로 가장 직관적인 방법인 동일 빈도 (equal frequency binning) 방법을 선택하였다. 의사결정나무를 적용한 매매규칙의 생성은 다음의 과정을 따른다. 트레이닝 데이터에 대한 수익률을 계산한다. 의사결정나무는 데이터의 분류기법 알고리즘이므로 기술적 지표를 기준을 수익률을 분류한다. 이때 의사결정나무의 의사결정마디 (decision node)는 각기 하나의 기술적 지표에 대한 조건이다. 분할을 반복적으로 하며 불순도가 가장 적은 의사결정나무를 선택한다. 이때 불순도의 측정은 엔트로피 지수 (entropy measure)를 사용한다. 반응변수  $m$ 개의 집단이

존재할 때 하나의 의사결정나무 A에 대한 엔트로피 지수는 다음과 같이 정의된다.

$$\text{entropy}(A) = - \sum_{k=1}^m p_k \log_2(p_k)$$

A : 의사결정나무의 분할노드

m : 분할노드 A에 속한 클래스의 수

$p_k$  : 분할노드 A에 있는 임의의 투플이 클래스에 속할 확률

선별된 기술적 지표들에 대해서 의사결정나무로 불순도가 낮은 분할값 (splitting value)을 찾아낸다.

이후에 두 가지의 방법으로 선별된 기술적 지표와 분류범위를 러프집합을 이용하여 매매규칙을 생성한다. 선별된 기술적 지표와 이산화 구간 혹은 분할값을 조합하여 리덕트 (reduct)를 생성한다. 이 리덕트는 러프집합의 기본 단위로 기술적 지표와 이산화 구간의 2가지 속성을 가지므로 2차원 테이블의 형태를 갖는다. 기술적 지표와 이산화 구간의 개수에 따라 조합 (combination)의 숫자가 결정되지만, 이들이 모두 수익을 내는 매매규칙으로 판단 할 수 없으므로, 트레이닝 데이터를 적용하여 그 유효성을 검증한 후, 유효성이 높은 리덕트 만을 선별한다. 유효성을 검증하는 방법은 트레이닝 데이터가 각각의 리덕트에 속하는 빈도를 측정하는 리덕트 정확성 (accuracy of approximation)을 측정하여 사용한다. 정확성이 높은 리덕트를 정규화된 매매규칙으로 사용한다.

### 3단계. 선물시장에서의 백테스팅을 통한 매매규칙의 성과분석

전 단계에서 생성된 매매규칙들에 대해서 성과분석을 실시한다. 성과 분석을 위한 선물 데이터는 매매규칙을 생성할 때 사용한 기간과 상이한 기간을 선택하여 중복되지 않는 다른 시장 환경에 대하여 성과를 분석 한다. 각기 매매규칙마다 개별적인 고유의 구조에서 포지션을 보유하는 것으로, 다른 매매규칙에 대한 영향 없이 독립적인 매매를 한다. 백테스트 데이터의 시계열이 시간에 따라 매매규칙에 적용되어 매매신호를 발생시킨다. 매매규칙이 발생시키는 매매신호는 매수, 매도의 단 두 가지로, 신호 발생 시 기계적으로 지연 없이 해당하는 선물 포지션을 보유한다. 다만 현실적인 매매환경을 구성하여 실제 선물 시장에서 발생하는 제약들을 반영하기 위해 매매 시 마다 슬리피지와 거래비용을 적용한다. 이는 매매규칙의 성과를 단조 감소시키지만, 매매횟수에 비례하여 성과를 감소시킨다. 이를 통해 최적의 매매성과에 대한 적절한 거래횟수 비율을 찾는 것을 기대한다.

## 4. 실증분석

본 연구에서 제안한 투자전략은 총 3단계로 구성되어 있다 (Figure 4.1). 이 연구에서 사용된 데이터는 국내 가장 거래가 활발하여 효율적 시장에 근접하는 한국 KOSPI 200 선물 시장에서의 데이터를 선택하였다. 데이터는 2002년 1월부터 2011년 12월까지 총 10년간의 데이터를 사용하였으며, 데이터의 시간단위는 거래 체결 데이터인 틱 단위 데이터를 사용하였다. 이 데이터는 한국증권전산인 코스콤 (<http://www.koscom.co.kr/>)에서 얻을 수 있다.

1 단계에서 수행하는 원시 데이터의 변환은 Table 4.1에서 표시된 14가지 시간 간격 데이터로 이루어진다. 각각의 매매신호 발생 간격이 다르므로 상이한 거래 패턴을 갖게 된다. 5분의 짧은 간격인 단기 매매에서부터 720분의 긴 간격 매매까지 다양한 시간간격에 대한 각각의 매매규칙을 생성하여 그 시간 간격에 대한 수익률 성과를 분석하였다.

이렇게 변환한 데이터를 크게 두 부분으로 나누어 트레이닝 데이터와 테스트 데이터로 사용하였다. 매매규칙 생성은 트레이닝 데이터에 대한 34개의 기술적 지표를 사용하였다. 이 중에서 트레이딩

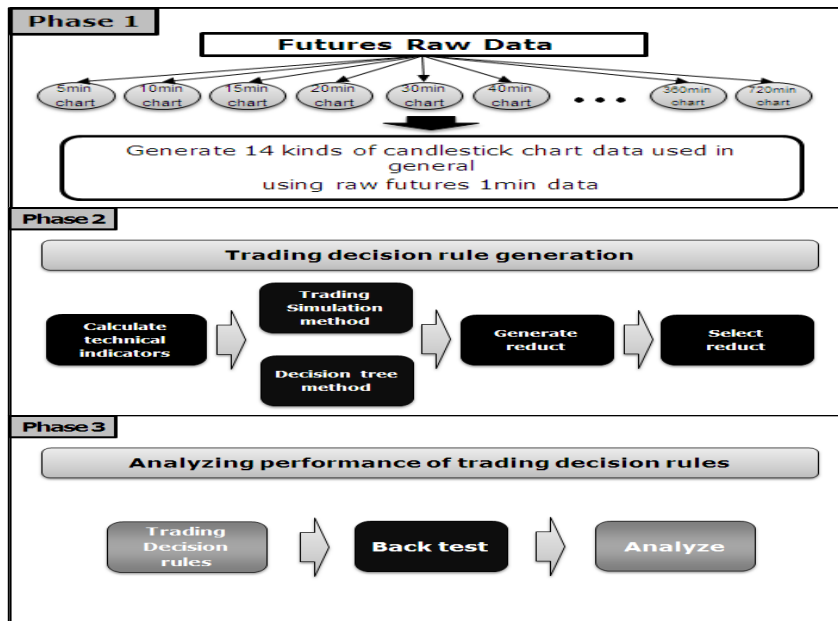


Figure 4.1 The proposed investment strategy of this study

Table 4.1 Classification of the transformed data

index	trading interval (min)	daily trading number
1	5	72
2	10	36
3	15	24
4	20	18
5	30	12
6	40	9
7	45	8
8	60	6
9	90	4
10	120	3
11	180	2
12	240	1.5
13	360	1
14	720	0.5

성과에 대한 적합도가 높은 소수의 기술적 지표만을 선택하여 매매규칙을 생성한다. 기술적 지표의 선택은 의사결정나무와 트레이딩 시뮬레이션의 두 가지 방법을 사용하여 전체 34개중 5개의 기술적 지표를 선택한다. 두 가지 선택방법에 대하여 트레이닝 구간 길이를 6달, 12달로 정의하여 매매규칙을 생성한다. 따라서 이렇게 생성된 매매규칙은 트레이닝 구간 길이, 매매규칙 생성 방법론, 매매 시간간격의 변수에 대한 다차원의 결과를 갖는다.

트레이딩 시뮬레이션 방법과 의사결정나무 방법에 의해 6개월 트레이닝으로 생성된 매매규칙의 테스트구간 시뮬레이션 결과는 Table 4.2와 Table 4.3에 나타내었다. 연중 선물지수의 변화와 비교하기 위해 초기 선물 매입포지션의 보유 수익률을 매수전략으로 표시하여 각 매매 시간 간격의 트레이딩

결과와 비교하였다. 60분 간격을 매매 시간 간격이 짧아질수록 수익률이 낮아지고, 길어질수록 수익률이 증대되었다. 이는 시간 간격의 조절로 시장 포트폴리오 전략보다 높은 수익률을 얻을 수 있음을 의미한다.

트레이딩 시뮬레이션 방법과 의사결정나무 방법의 매매규칙 성과를 비교하여 보면 수익률에서는 의사결정나무 방법이 약간 더 낮은 성과를 보였으나, 편차가 감소했으므로 명확하게 성과가 감소되었다고 보기는 어렵다. 의사결정나무 방법이 리스크 (편차)가 감소하는 만큼 수익률이 감소한 것으로, 높은 리스크에 높은 수익이 따른다는 법칙 (high risk high return) 을 만족시키는 것을 확인할 수 있다.

**Table 4.2** Yearly yield of 6 month training rules in trading simulation

trading interval	return rate (%)					
	2006	2007	2008	2009	mean	sd
long	6.399	10.267	-25.552	19.603	2.679	19.620
5	-17.136	-34.552	23.240	-47.647	-19.024	30.823
10	-29.533	-32.583	-3.122	-43.835	-27.268	17.232
15	-37.169	-36.944	-27.637	-43.090	-36.210	6.384
20	-19.562	-9.534	-24.350	-12.966	-16.603	6.632
30	-7.858	8.414	-25.693	11.377	-3.440	17.076
40	-2.817	8.853	-23.056	12.562	-1.115	16.028
45	-1.846	9.679	-25.845	16.587	-0.356	18.616
60	1.250	17.838	-33.535	24.253	2.451	25.875
90	2.517	19.772	-34.760	28.727	4.064	28.075
120	4.788	20.792	-31.381	27.287	5.372	26.262
180	4.397	16.434	-27.430	25.873	4.818	23.226
240	4.857	20.601	-27.796	27.537	6.300	24.631
360	4.106	13.152	-20.684	23.116	4.922	18.754
720	5.401	17.363	-26.196	29.097	6.416	23.797

**Table 4.3** Yearly yield of 6 month training rules in decision tree

trading interval	return rate (%)					
	2006	2007	2008	2009	mean	sd
long	6.399	10.267	-25.552	19.603	2.679	19.620
5	-50.086	-33.340	28.304	-47.268	-25.598	36.673
10	-30.099	-38.970	-0.301	-42.299	-27.917	19.117
15	-26.743	-32.261	-10.140	-27.332	-24.119	9.642
20	-22.227	-12.358	-29.339	-13.106	-19.257	8.081
30	-10.275	3.730	-27.289	8.517	-6.329	16.088
40	-2.973	9.661	-22.484	11.687	-1.027	15.707
45	0.059	13.033	-29.515	21.217	1.199	22.252
60	2.359	18.260	-35.719	25.480	2.595	27.308
90	2.549	19.982	-34.153	26.766	3.786	27.272
120	2.711	17.837	-25.553	24.074	4.767	22.114
180	3.871	17.474	-24.684	23.934	5.149	21.575
240	3.699	17.373	-24.523	26.213	5.691	22.170
360	4.460	12.483	-22.573	22.735	4.276	19.399
720	5.720	15.601	-27.845	26.724	5.050	23.548

앞의 6개월 트레이닝 기간 매매규칙과 비교하여 Table 4.4와 Table 4.5에서는 12개월 트레이닝 기간을 거쳐 생성된 매매규칙의 동일 구간 테스트 결과를 나타내었다. 12개월 트레이닝 기간의 매매규칙도 마찬가지로 트레이딩 시뮬레이션 방법과 의사결정나무 방법을 비교하면, 6개월 트레이닝

매매규칙과 동일하게 의사결정나무 방법을 사용한 경우가 다른 경우보다 수익률과 편차가 낮아지는 것을 볼 수 있다. 이 경우에도 마찬가지로 수익률과 리스크가 양의 상관관계를 갖는다는 점을 입증하는 것으로 판단할 수 있다. 다만, 12개월의 트레이닝 매매규칙이 6개월 트레이닝 매매규칙보다 더 높은 수익률과 편차를 보이는 것으로, 트레이딩 기간의 길이가 트레이딩의 성과에 미치는 영향을 비교할 수 있다.

**Table 4.4** Yearly yield of 12 month training rules in trading simulation

trading interval	return rate (%)				mean	sd
	2006	2007	2008	2009		
long	6.399	10.267	-25.552	19.603	2.679	19.620
5	-8.252	-27.347	38.260	-41.758	-9.774	34.840
10	-25.216	-29.200	13.371	-37.426	-19.618	22.572
15	-21.521	-29.563	10.125	-35.337	-19.074	20.274
20	-19.020	-11.656	-17.788	-19.799	-17.066	3.700
30	-13.454	4.616	-17.643	2.626	-5.964	11.228
40	-7.216	9.444	-24.897	13.228	-2.360	17.453
45	-6.200	8.695	-26.754	14.959	-2.325	18.547
60	0.634	15.333	-30.529	23.207	2.161	23.716
90	1.999	22.861	-40.763	33.593	4.423	32.856
120	3.312	23.300	-36.750	33.423	5.821	31.016
180	5.264	23.611	-35.642	32.678	6.478	30.307
240	6.292	26.641	-38.485	35.150	7.400	32.898
360	5.923	20.234	-30.878	31.237	6.629	27.067
720	6.684	21.720	-34.908	34.056	6.888	30.028

**Table 4.5** Yearly yield of 12 month training rules in decision tree

trading Interval	return rate (%)				mean	sd
	2006	2007	2008	2009		
long	6.399	10.267	-25.552	19.603	2.679	19.620
5	-7.671	-26.564	39.348	-40.348	-8.809	34.787
10	-21.479	-33.697	13.110	-41.960	-21.007	24.250
15	-19.061	-31.413	10.810	-35.090	-18.688	20.826
20	-24.233	-19.040	-9.975	-26.358	-19.902	7.297
30	-14.251	-1.404	-19.755	1.094	-8.579	10.036
40	-6.727	8.136	-19.907	11.201	-1.824	14.375
45	-2.630	11.447	-27.925	14.737	-1.093	19.409
60	-1.493	17.413	-33.099	21.962	1.196	25.017
90	1.247	21.174	-36.245	29.449	3.906	29.268
120	3.438	25.329	-37.348	33.542	6.240	31.715
180	4.816	23.815	-35.642	32.741	6.432	30.371
240	5.227	24.447	-36.680	33.545	6.635	31.196
360	5.784	21.227	-31.702	31.441	6.687	27.681
720	5.643	22.206	-31.824	30.303	6.582	27.585

위 결과에서 잦은 빈도의 매매 규칙은 저 빈도의 매매 규칙보다 수익이 크게 악화 되는 경향을 보인다. 이는 매 거래 시마다 거래비용과 슬리피지의 손실을 보기 때문이다. 이 손실은 선물 포지션의 진입, 청산시마다 발생한다. 따라서 거래 횟수에 비례하여 손실이 증가하므로, 매매로 인한 이익이 손실의 폭보다 큰 시간 간격 매매 규칙을 찾아야 한다. 각 매매규칙을 생성하는 경우 156개의 매매규칙이 생성되는데 이 중에 가장 수익률이 좋은 매매규칙의 거래횟수와 평균 거래횟수를 비교한 것이 Table 4.6와 Table 4.7에서 보여준다.



**Table 4.6** Transaction number of 6 month training rules

	average	best case	ratio (%)
5	206.163	192.750	93.494
10	403.202	44.000	10.913
15	566.389	213.250	37.651
20	247.950	106.750	43.053
30	103.438	23.750	22.961
40	60.029	56.500	94.121
45	49.511	126.000	254.488
60	18.146	10.000	55.109
90	7.160	47.750	666.893
120	3.169	0.750	23.665
180	1.642	16.750	1019.97
240	1.315	4.250	323.239
360	0.707	0.750	106.071
720	0.719	0.750	104.348

**Table 4.7** Transaction number of 12 month training rules

	average	best case	ratio (%)
5	60.614	1.000	1.650
10	286.298	83.750	29.253
15	267.329	106.750	39.932
20	248.189	93.250	37.572
30	141.345	111.500	78.885
40	80.721	28.000	34.687
45	63.256	79.500	125.679
60	31.304	30.000	95.833
90	9.824	9.750	99.250
120	5.593	9.750	174.327
180	1.295	12.750	984.653
240	3.162	9.750	308.363
360	0.994	5.750	578.710
720	1.056	0.750	71.017

각기 156개의 매매규칙 중에서 최고의 수익을 내는 단일 매매규칙의 거래횟수는 156개 평균 거래 횟수보다 보통 적게 나타난다. 특히, 고빈도의 매매를 하게 되는 짧은 거래 간격의 매매규칙에서는 평균 거래횟수에 비해 좋은 수익을 나타낼 때의 거래횟수가 크게 줄어든다. 따라서 특정 빈도 이상의 매매규칙에 대해서, 슬리피지와 거래비용의 영향으로 수익이 크게 감소하는 원인을 생각할 수 있다.

Figure 4.2와 Figure 4.3는 각 각의 매매규칙에 대한 샤프지수를 보여준다. 수익률이 높고, 편차가 낮을수록 샤프지수가 높아지므로, 짧은 매매 간격에 대해서는 샤프지수가 낮으며 매매 간격이 길어질수록 샤프지수도 증가한다. 60분미만의 시간 간격 구간에서는 음의 수익률로 인해 샤프지수도 음수로 결정된다. 편차가 가장 적은 20분 시간 간격에서 샤프지수의 변화율이 가장 크며, 수익률이 기본 보유 전략과 근접하는 60분 시간 간격에서 샤프지수의 값이 0에 수렴하게 된다. 이는 기본 보유 전략이 무위험 이자율에 상당하는 수익률을 얻는다고 할 때, 60분 시간 간격의 매매규칙이 이와 동일한 수준의 수익률을 낸다는 것을 의미한다. 이후 더 긴 시간 간격의 수익률이 높아지며 샤프지수도 양으로 전환되지만, 변화율이 낮아져 크게 수익이 좋아지지는 않는다. 샤프지수의 변화율이 0에 수렴하게 되어 수익률이 정체된다.

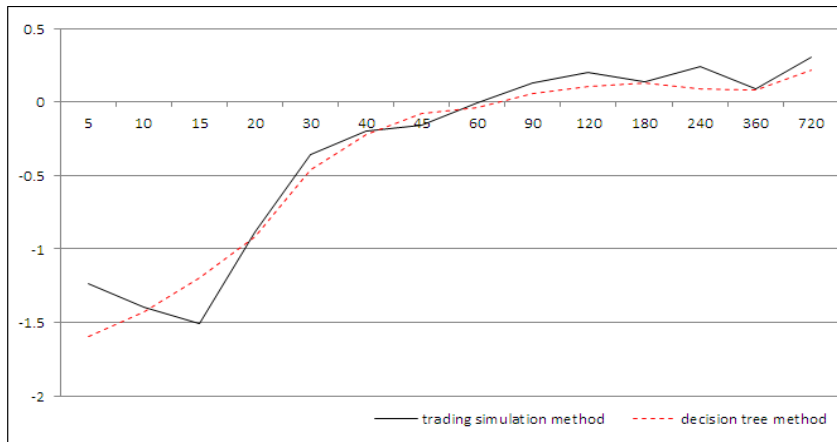


Figure 4.2 Sharpe ratio of 6 month training rules

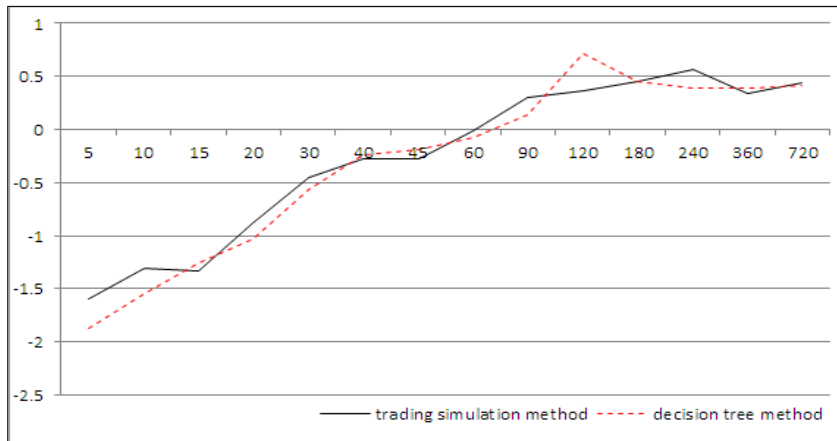


Figure 4.3 Sharpe ratio of 12 month training rules

## 5. 결론

KOSPI 200 선물시장은 국내에서 가장 활발히 거래되는 금융시장 중 하나로써, 세계에서 가장 거래량이 많은 금융시장 중 하나이다. 파생상품인 선물의 특성상 현물에 대한 헤지 수요와 투자 자산, 단기 매매를 위한 투기 자산의 역할을 하며, 바스켓으로 운용하는 현물에 비해 인덱스 상품의 특성을 갖기 때문에 기초자산인 현물보다도 더 선호되는 금융자산이다. 높은 거래량은 시장참여자의 거래비용을 낮추고, 이론가와 시장가의 괴리를 줄이며, 시장가격에 왜곡이 생기는 경우 빠르게 정상가격으로 회복시키는 효과를 갖는다. 따라서 KOSPI 200 선물시장은 국내에서 가장 효율적이고 유동적인 금융시장이다. 즉, 노이즈나 왜곡 없이 시장에서 발생시키는 시그널에 대한 시스템 트레이딩에 가장 적합한 시장이라 할 수 있다. 따라서 본 연구는 러프집합과 의사결정나무를 이용하여 선물시장에서의 시스템 트레이딩을 위한 효율적인 매매규칙을 찾고자 하였다. 제 4절에서 보여준 실험 결과로부터

본 연구에서 제안한 매매규칙을 사용한 시스템 트레이딩은 다양한 패턴을 발생시키는 선물시장에서 투자자들에게 비교적 안정적이면서 높은 수익률을 낼 수 있도록 도움을 주는 거래 전략을 제공하기에 충분하다고 판단할 수 있었다. 본 연구에서는 14가지의 매매 시간 간격의 선물 데이터를 사용함으로써 제안한 투자전략의 매매규칙을 통해 다양한 선물 매매 간격에 제약 받지 않을 수 있는 유용한 투자 전략을 제시하였다. 이를 통해 본 연구 결과가 많은 투자자들에게 투자 결정을 함에 있어서 중요한 정보를 제공할 수 있을 것으로 기대해 본다.

본 연구에서는 러프집합과 의사결정나무를 이용하여 투자전략을 구축하였다. 하지만 금융시장에서 획득되는 비선형적인 데이터들에 대해 최근 다양한 인공지능 알고리즘 기법들이 금융시장 분석 기법으로 적합하다는 연구가 진행 중이다. 인공신경망 (artificial neural network; ANN) 과 유전자 알고리즘 (genetic algorithm) 과 같은 다양한 분류 알고리즘을 사용한다면 더 좋은 결과가 나타날 수 있기 때문에 이러한 다양한 분석을 통해 추후 연구에서는 선물시장에서의 수익률을 극대화 할 수 있는 더 나은 투자 전략 시스템이 나올 수 있기를 기대한다.

### 참고문헌

- Arie, B. D. (2008). Rule effectiveness in rule-based systems: A credit scoring case study. *Expert Systems with Applications*, **34**, 2783-2788.
- Byun, H. W., Song, C. W., Han, S. K., Lee, T. K. and Oh, K. J. (2009). Using genetic algorithms to develop volatility index-assisted hierarchical portfolio optimization. *Journal of the Korean Data & Information Science Society*, **20**, 467-478.
- Chavarnakul, T. and Enke, D. (2008). Intelligent technical analysis based equivolume charting for stock trading using neural networks. *Expert Systems with Applications*, **34**, 1004-1017.
- Chavarnakul, T. and Enke, D. (2009). A hybrid stock trading system for intelligent technical analysis-based equivolume charting. *Neurocomputing*, **72**, 3517-3528.
- Chris, F., Andros, G. and Alexandros, K. (2011). Trading frequency and asset pricing on the London Stock Exchange: Evidence from a new price impact ratio. *Journal of Banking & Finance*, **35**, 3335-3350.
- Dymova, L., Sevastianov, P. and Bartosiewicz, P. (2010). A new approach to the rule-base evidential reasoning: Stock trading expert system application. *Expert Systems with Applications*, **37**, 5564-5576.
- Gencay, R. (1999). Linear, non-linear and essential foreign exchange rate prediction with simple technical trading rules. *Journal of International Economics*, **47**, 91-107.
- Hong, J. S. and Kwon, T. W. (2010). Distribution fitting for the rate of return and value at risk. *Journal of the Korean Data & Information Science Society*, **21**, 219-229.
- Horton, M. J. (2009). Stars, crows, and doji : The use of candlesticks in stock selection. *The Quarterly Review of Economics and Finance*, **49**, 283-294.
- Ito, A. (1999). Profits on technical trading rules and time-varying expected returns: Evidence from Pacific-Basin equity markets. *Pacific-Basin Finance Journal*, **7**, 283-330.
- Izumi, K., Toriumi, F. and Matsui, H. (2009). Evaluation of automated-trading strategies using an artificial market. *Neurocomputing*, **72**, 3469-3476.
- Kho, B. C. (1996). Time-varying risk premia, volatility, and technical trading rule profits: Evidence from foreign currency futures markets. *Journal of Financial Economics*, **41**, 249-290.
- Kim, M. S. and Oh, K. J. (2011). Using rough set to support arbitrage box spread strategies in KOSPI 200 option markets. *Journal of the Korean Data & Information Science Society*, **22**, 37-47.
- Lee, J. H. and Jo, G. S. (1999). Expert system for predicting stock market timing using a candlestick chart. *Expert Systems with Applications*, **16**, 357-364.
- Lee, S. J. and Oh, K. J. (2011). Finding the optimal frequency for trade and development of system trading strategies in futures market using dynamic time warping. *Journal of the Korean Data & Information Science Society*, **22**, 255-267.
- Marc, B. (2004). Khiops: A statistical discretization method of continuous attributes. *Machine Learning*, **55**, 53-69.
- Marshall, B. R., Young, M. R. and Rose, L. C. (2006). Candlestick technical trading strategies: Can they create value for investors? *Journal of Banking & Finance*, **30**, 2303-2323.

- Murphy, J. J. (1999). *Technical analysis of the financial markets*, New York Institute of Finance, Paramus, NJ.
- Neely, C. J. (2002). The temporal pattern of trading rule returns and exchange rate intervention: intervention does not generate technical trading profits. *Journal of International Economics*, **58**, 211-232.
- Oh, K. J., Kim, T. Y., Jung, K. W. and Kim, C. H. (2011). Stock market stability index via linear and neural network autoregressive model. *Journal of the Korean Data & Information Science Society*, **22**, 355-351.
- Park, B. J. (2011). Herd behavior and volatility in financial markets. *Journal of the Korean Data & Information Science Society*, **22**, 1199-1215.
- Park, I. S., Han, J. T., Kang, S. B. and Ji, J. H. (2010). Developing the predictive model for stomach cancer using data mining. *Journal of the Korean Data & Information Science Society*, **21**, 1253-1261.
- Pawlak, Z. (1991). *Rough sets - Theoretical aspects of reasoning about data*, Kluwer Academic Publishers, Dordrecht, Boston, London.
- Pawlak, Z. (1997). Rough set approach to knowledge-based decision support. *European Journal of Operational Research*, **99**, 48-57.
- Pawlak, Z. and Munakata, T. (1996). Rough control, application of rough set theory to control. *Proceedings of the 4th European Congress on Intelligent Techniques and Soft Computing*, Germany, 209-218.
- Quinlan, J. R. (1986). Induction of decision tree. *Machine Learning*, **1**, 81-106.
- Sharpe, W. F. (1994). The Sharpe ratio. *Journal of Portfolio Management*, **21**, 49-58.
- Shen, L. and Loh, H. T. (2004). Applying rough sets to market timing decisions. *Decision Support Systems*, **37**, 583-597.
- Shin, Y. K. (2006). An empirical study on stock trading value of each investor type in the Korean stock market. *Journal of the Korean Data & Information Science Society*, **17**, 1099-1106.
- Slowinski, R. and Stefanowski, J. (1994). *Rough classification with valued closeness relation*, Springer-Verlag, Berlin, 482-488.
- Song, C. W. and Oh, K. J., (2009). Study of validation process according to various option strategies in a KOSPI 200 options market. *Journal of the Korean Data & Information Science Society*, **20**, 1061-1073.
- Tan, Z., Quek, C. and Cheng, P. Y. K. (2011). Stock trading with cycles: A financial application of ANFIS and reinforcement learning. *Expert Systems with Applications*, **38**, 4741-4755.
- Taylor, M. P. (1992). The use of technical analysis in the foreign exchange market. *Journal of International Money and Finance*, **11**, 304-314.
- Wei, J. M., Wang, S. Q., Wang, M. Y., You, J. P. and Liu, D. Y. (2007). Rough set based approach for inducing decision trees. *Knowledge-Based Systems*, **20**, 695-702.

## Using rough set to develop the optimization strategy of evolving time-division trading in the futures market

Hyun Ho Kim<sup>1</sup> · Kyong Joo Oh<sup>2</sup>

<sup>12</sup>Department of Information and Industrial Engineering, Yonsei University

Received 31 May 2012, revised 25 June 2012, accepted 27 July 2012

### Abstract

This paper proposes to develop system trading strategy using rough set, decision tree in futures market. While there is a great deal of literature about the analysis of data mining, there is relatively little work on developing trading strategies in futures markets. There are three objectives in this paper. The first objective is to analysis performance of decision tree in rule-based system trading. The second objective is to find proper profitable trading interval. The last objective is to find optimized training period of trading rule training. The results of this study show that proposed model is useful trading strategy in foreign exchange market and can be desirable solution which gives lots of investors an important investment information.

*Keywords:* Decision tree, futures market, rough set, system trading, trading optimization.

---

<sup>1</sup> Graduate student, Department of Information and Industrial Engineering, Yonsei University, Seoul 120-749, Korea.

<sup>2</sup> Corresponding author: Associate professor, Department of Information and Industrial Engineering, Yonsei University, Seoul 120-749, Korea. E-mail: johanoh@yonsei.ac.kr