Journal of the Korean Data & Information Science Society
2014, **25**(2), 281–292

http://dx.doi.org/10.7465/jkdi.2014.25.2.281
한국데이터정보과학회지

선물시장에서 러프집합 기반의 유전자 알고리즘을 이용한 최적화 거래전략 개발

정승환[1] · 오경주[2]

[12]연세대학교 정보산업공학과

접수 2013년 12월 2일, 수정 2013년 12월 26일, 게재확정 2014년 1월 27일

요 약

최근 알고리즘 트레이딩에 대한 관심이 높아지면서, 인공지능 방법론을 이용한 매매 전략 구축에 관련된 연구들이 활발하게 진행되고 있다. 하지만 복수의 인공지능 방법론을 융합하여 매매 전략 개발에 이용한 사례는 아직 많지 않다. 본 연구는 주가지수선물시장을 바탕으로 인공지능 방법론 중 하나인 러프집합 이론을 적용하여 알고리즘 트레이딩 매매전략을 개발한다. 특히 유전자 알고리즘을 도입하여 생성된 매매전략을 현재시장상황에 최고의 수익률을 보일 수 있도록 최적화한다. 실증분석으로는 2009년부터 2012년까지 4년간의 매매수익률을 분석한 결과 매수 후 보유 전략과 비교하여 우수한 성과를 보였다.

주요용어: 러프집합, 선물시장, 알고리즘 트레이딩, 유전자 알고리즘.

1. 서론

KOSPI200 선물시장의 변동성이 증가함에 따라 시장 방향성의 예측이 중요해지고 있으며, 학계에서도 이에 대한 관심이 집중되고 있다. 장기적, 단기적 관점에서 주식시장의 방향을 예측하기 위해 각종 이론과 시장정보를 이용해 기본적 분석, 기술적 분석, 통계적 분석 등 다양한 방식으로 연구되고 있다 (Lo 등, 2000). 통계적 분석 방법론으로 판별분석이나 로지스틱 회귀분석 등을 이용해서 주식의 변화를 분석하는 연구가 진행되어 왔으며, 이러한 연구가 최근에는 인공지능 방법론을 이용한 고차원 분석기법으로 발전해왔다. 인공 신경망 (artificial neural network), 유전자 알고리즘 (genetic algorithm), 퍼지 이론 (fuzzy theory), 사례기반 추론 (case based reasoning) 등 다양한 인공지능 방법론을 금융시계열 데이터에 적용하여 주식시장을 예측하기 위한 알고리즘 트레이딩에 대해 연구되고 있다 (Kim 등, 2000; Dong 등, 2009; Kang 등, 2013). 알고리즘 트레이딩 (algorithm trading)은 인간의 주관적인 판단을 배제하고 일정한 규칙에 따라 기계적으로 매매하여 투자수익을 얻는 거래 시스템을 의미한다. 인공지능 방법론을 활용한 알고리즘 트레이딩은 기계학습 (machine learning)방식으로 시스템의 성능을 향상시키는 것이 특징이다. 다양한 방법론을 이용한 연구가 선행되어왔는데, Nunes-Letamendia (2007)은 기술적 분석을 이용한 알고리즘 트레이딩의 최적화에 유전자 알고리즘을 접목시켰으며, Chavarnakul와 Enke (2009)는 인공신경망, 퍼지, 유전자 알고리즘을 활용한 임계치 기반의 알고리즘 트레이딩을 제안하였다. Bao와 Yang (2008)은 주식정보의 고차원적 표현을 확률모형과 결합하는 인공지능 트레이딩

[1] (120-749) 서울특별시 서대문구 신촌동 134번지, 연세대학교 정보산업공학과, 석사과정.
[2] 교신저자: (120-749) 서울특별시 서대문구 신촌동 134번지, 연세대학교 정보산업공학과, 부교수.
E-mail: johanoh@yonsei.ac.kr

시스템을 개발하였다. 또한 Zarandi 등 (2009)과 Chang와 Liu (2008)는 주식 가격 분석을 위한 전문가 시스템을 구축하기 위해 퍼지 이론를 도입하여 규칙 기반의 의사 결정 지원 시스템을 제안하였다.

본 연구에서는 집합 형성을 통한 의사결정 규칙 방법론 중 하나인 러프집합 이론 (rough-set theory)을 이용하여 주가지수선물시장의 방향성을 예측하는 거래모형을 제안한다. 주가지수선물시장의 역사적 데이터에 러프집합 이론을 접목하여 시장의 방향성을 상승 또는 하락으로 예측할 수 있는 거래규칙을 생성한다. 러프집합 이론을 이용한 분석은 투자자가 매매하고자 하는 시점의 과거 일정 구간 동안의 데이터를 활용하여 현재시장에서 예측력을 가지고 있는 거래규칙을 생성한다. 투자자는 러프집합이론으로 생성된 거래규칙을 바탕으로 투자의견 전문가 시스템 (expert system)으로 활용할 수 있으며 거래규칙만으로 매매전략을 수립하는 알고리즘 트레이딩 기법으로도 활용할 수 있다. 또한 본 연구에서는 러프집합 이론에 사용되는 다양한 모수 (parameter)들을 유전자 알고리즘 기법을 사용하여 최적화하였다. 시스템 매매전략의 투자수익률을 극대화하는 것을 유전자 알고리즘의 목적함수로 설정하여 거래규칙의 효율성을 극대화하기 위하여 노력하였다.

본 논문은 다음과 같이 구성되어 있다. 2절에서는 제안된 모형에 사용된 방법론에 대한 선행연구를 기술하였으며, 3절에서는 본 연구에서 제안하는 알고리즘 트레이딩 전략에 대해 자세히 설명하고, 4절에서는 제안된 전략에 대한 실증분석을 하였으며, 마지막으로 결론에서는 본 연구의 기대효과 및 향후 연구방향에 대해 기술하였다.

2. 선행 연구

2.1. 러프집합 이론

러프집합은 Pawlak에 의해 제안된 것으로 불확실성 (uncertainty), 모호함 (vagueness), 부정확성 (imprecision)을 가지고 있는 데이터집합에서 일관성을 가지는 규칙을 찾아내기 위한 수치해석적 접근방법이다 (Pawlak, 1997). 러프집합 이론은 모든 개체들은 항상 집합이 형성될 수 있다고 가정하고 일정한 정보들을 바탕으로 개체들의 동질성 관계 (indiscernibility relationship)을 찾아내는 것이 목적이다. 동질성을 가진 집합을 기본집합 (elementary set)이라고 하며 동질성관계로 분류되지 못한 집합을 러프집합 (rough set)이라고 정의한다.

러프집합은 하한 근사영역 (lower approximation)과 상한 근사영역 (upper approximation)으로 설명되는데, 동질성이 확실하게 분류되는 부분은 하한 근사 영역, 그렇지 않은 부분을 상한 근사영역이라고 정의한다. 근사 영역간의 경계영역 (boundary region)을 계산할 수 있으며, 경계영역의 집합들을 하한 근사영역의 정보들을 바탕으로 근사하여 동질성 관계를 가지는 집합들을 찾아낼 수 있다.

러프집합과 반대되는 개념인 정확집합 (exact set)을 통해 러프집합 이론에 대해 설명할 수 있다. 정확집합 또는 일반집합 (crisp set)이란, 집합 내의 원소 모두가 확실하게 정의 가능한 집합을 의미한다. 반면, 집합 내의 원소 중 주어진 조건속성들로 결정속성을 명확하게 정의할 수 없는 집합을 러프집합 (rough set)이라고 한다. Table 2.1은 A, B의 2가지 조건속성과 C의 결정속성을 가지고 있는 집합의 예이다. 이때, 1∼4의 원소들을 전체 집합이라고 하면, 조건속성들로 결정속성 C가 명확하게 정의되기 때문에 정확집합이라고 할 수 있다. 하지만 1∼6의 원소들을 전체집합이라고 하면 A, B가 각각 3, 1일 때, 결정속성 C가 3 또는 4로 결정되기 때문에 결정할 수 없어 러프집합으로 분류된다. 러프집합 이론은 이러한 러프집합내의 부정확한 결정속성을 상한근사영역으로 정의하고 근사하여 분류한다. 집합내에 동일한 조건속성을 가지는 원소들이 가지는 결정속성의 수를 합산하여 가장 높게 지지되는 결정속성을 해당 조건속성의 결정값으로 근사한다.

Table 2.1 Example of rough set theory

ID	A	B	C
1	1	1	**1**
2	1	2	**1**
3	2	1	**2**
4	2	2	**3**
5	3	1	**3**
6	3	1	**4**

2.2. 유전자 알고리즘

유전자 알고리즘은 생물의 진화원리로부터 착안된 알고리즘으로 확률적 탐색이나 학습 및 최적화를 위한 방법론이다. 멘델 (Mendel)의 유전법칙과 찰스 다윈 (Charles Darwin)의 적자생존의 원리를 적용하여 생물이 진화하는 방식대로 개체들을 진화 또는 최적화한다. 유전자 알고리즘은 개체군 (chromosome pool)을 구성하고, 각 염색체 (chromosome)에 대해 병렬적으로 탐색이 이루어지기 때문에 탐색의 방향이나 초기값의 설정에 과도하게 의존하지 않고 세대 (generation)의 진화에 따라 확률적으로 진화한다는 특징을 가지고 있다. 유전자 알고리즘은 크게 4가지의 프로세스로 진행된다. 첫째, 초기의 개체군을 형성하는 염색체들을 생성하고 이를 모집단이라고 정의한다. 둘째, 각 염색체들의 문제해결에 대한 적합성을 판단하기위해 적합도 함수 (fitness function)을 통해 적합도를 평가한다. 셋째, 평가된 염색체들의 적합도를 비교하여 정해진 세대의 규모를 초과하면서 적합도가 낮은 염색체들은 개체군에서 제외시킨다. 넷째, 선택 (selection), 교배 (crossover), 돌연변이 (mutation)을 통해 각 염색체들에 대한 유전자를 조작하여 다음 세대의 개체군을 형성한다. 이러한 과정을 통해 가장 최적의 적합 정도를 나타낸 개체를 유사 전역 최적해로 선택하게 된다.

유전자 알고리즘을 금융에 적용시킨 선행 연구로는 Kim과 Han (2000)은 인공신경망 방법론에 유전자 알고리즘을 접목하여 주가지수를 예측하는 모델을 제안하였고, Kuo 등 (2001)은 유전자 알고리즘을 기반으로 하는 퍼지 인공신경망 모델을 이용하여 주식 거래 지원시스템을 제안하였다. Kim과 Ahn (2010)이 서포터 벡터 기계 (support vector machine)에 유전자 알고리즘을 활용하여 등락을 예측하는 지능형 트레이딩 시스템을 개발하였다.

3. 연구 모형

본 연구에서 제안하는 모형은 크게 거래규칙 생성 단계, 거래 시뮬레이션 및 수익률 측정 단계, 유전자 알고리즘을 적용한 거래규칙 최적화 단계의 총 3단계로 구성된다. 첫 번째, 거래규칙 생성 단계에서는 입력변수로 사용되는 기술적지표를 가공하여 러프집합 분석을 통해 거래규칙을 생성하고 유의미한 거래규칙을 선별한다. 두 번째, 거래 시뮬레이션 및 수익률 측정 단계에서는 완성된 거래규칙을 시뮬레이션 구간에 적용하여 거래규칙 기반의 트레이딩을 진행하고 수익률로 거래규칙의 성과를 측정한다. 마지막으로 측정된 수익률을 적합도함수로 하는 유전자 알고리즘 최적화를 진행하여 해당 구간에서 최적화된 거래규칙을 찾아낸다. 최적화의 요인으로는 입력변수의 변수선택과 이산화 요인의 결정이 반영된다.

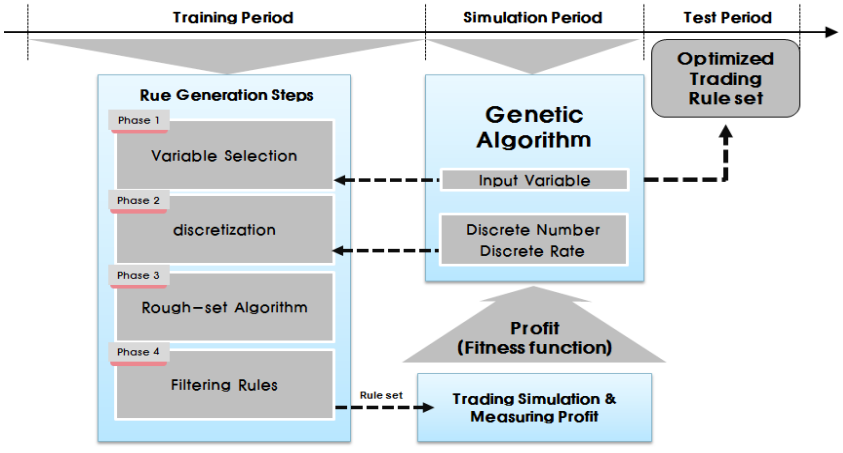

Figure 3.1 Diagram of purposed trading system

3.1. 기술적지표 및 이산화 과정

연구 모형의 입력변수로 주가지수 선물시장에 대한 기술적지표를 해당 트레이닝 구간의 선물가격정보를 바탕으로 생성하여 사용한다. 기술적지표에는 추세추종형 (trend following) 지표와 수익추종형 (profit following) 지표가 있다. 추세추종형 지표는 가격의 추세의 강도나 방향을 지표화하여 투자보조정보를 제공하는 지표를 의미하며, 수익추종형 지표는 거래강도나 이동평균 (moving average) 등을 이용하여 수익을 위한 매매시그널을 제공하는 지표를 의미한다 (Simon, 2011). 본 연구에서는 추세를 반영하면서 수익추종형지표의 매매시그널을 이용하기 위하여 항상 거래규칙에 두 유형의 지표가 모두 포함되도록 변수를 선택하였다. 또한, 기술적지표는 오실레이터 (oscillator)형과 비오실레이터형으로 구분할 수 있는데, 오실레이터형 지표는 지표를 생성하는 구간에 상관없이 일정한 범위 내에서 값이 유지되는 반면에 비오실레이터형 지표는 지표 생성 시작지점으로부터 값이 누적되어 증가 혹은 감소하는 특성을 가지고 있다. 본 연구에서는 거래규칙 생성구간과 적용구간이 다르기 때문에 구간에 독립적으로 일정범위 내에서 값이 생성되는 오실레이터형 지표만을 사용하였다.

본 연구에서 러프집합 알고리즘을 통해 얻고자 하는 결과가 조건속성과 결정속성으로 이루어진 거래규칙이기 때문에 연속형의 시계열데이터인 지표값을 러프집합 분석의 입력변수로 사용하기 위해 이산화 (discretization)하여 범주형 자료로 변환한다 (Marc, 2004). 러프집합 분석은 불확실하게 분류되어있는 상한 근사공간에 해당하는 집합을 분류하는 것이 목적인데, 조건속성을 연속형 변수를 사용하게 되면 분류해야하는 집합을 구성할 수 없으므로 범주형 변수로 이산화하는 과정이 필요하다.

이산화과정에서 고려해야 하는 요인은 크게 2가지가 있는데, 이산화 분위수와 이산화 비율이며 (Dai 등, 2000) 유전자 알고리즘의 모수로 사용된다. 이산화 분위수는 해당 기술적지표를 몇 구간으로 이산화 할 것인지를 의미하며, 이산화 비율이란 각 이산화 구간의 범위 비율을 상대적으로 나타낸다. Figure 3.2는 기술적지표 중 하나인 MFI (money flow index)의 이산화에 대한 예시이다. MFI는 최근 n일 간 주가의 상승간격과 하락간격의 비율을 백분율로 나타낸 기술적 지표로, 자세한 식은 Table 4.1에 기술하였다. MFI 지표를 4등분 동등 간격으로 이산화하면 0~25, 25~50, 50~75, 75~100의 4가지 범주를 가지는 (a)와 같은 형태의 범주형 변수가 된다. 반면 (b)는 범주를 3분위로, 범주의 간격을 비균등으로 나누어 범주형 변수를 생성한 결과이다. 이와 같이, 동일한 연속형 변수를 이산화하더라도 분위수와 비율에 따라 다른 범주형 변수가 생성되기 때문에 이 두 변수는 이산화에서 중요하게 작용하는 모수이다. 본 연구에서는 거래규칙을 최적화하는 조건속성, 즉 범주형 변수를 찾기 위해 분위수와 이산화 비율을 유전자알고리즘의 최적화 모수로 선택하였다. 그러므로 유전자 알고리즘은 최적화를 통해 최적의 거래결과를 가져오는 범주형 변수를 찾게 된다.

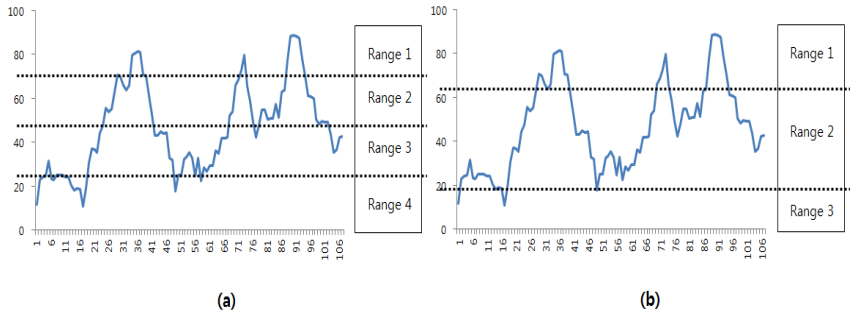

Figure 3.2 Example of discretization

3.2. 러프집합을 이용한 거래규칙 생성

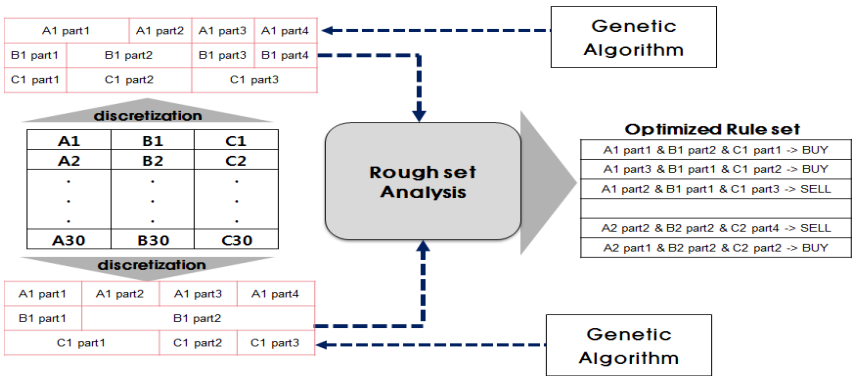

Figure 3.3 Architecture of trading strategy rule generation phase

이산화된 기술적지표를 입력변수로 사용하여 러프집합 분석을 수행하면 본 연구에서 제안하는 알고리즘 트레이딩을 위한 거래규칙을 생성할 수 있다. 러프집합 분석은 집합 내의 원소가 모두 정의 가능한 집합인 정확집합과 정확히 분류 될 수 없는 집합인 부정확집합, 즉 러프집합으로 구성되어 있다. 정확집합들을 기반으로 부정확한 데이터를 분류하고 처리하고자 하는 것이 러프집합 분석의 목적이다. 본 연구에서는 범주형 변수로 이산화된 기술적지표에서 3개의 지표를 유전자 알고리즘으로 변수선택하여 조건속성으로 사용하였고, 그에 대한 결정속성은 선물종가의 상승 또는 하락이다. 즉, 기술적지표가 어떠한 범주에 속해 있을 때, 선물종가를 상승 또는 하락으로 예측할 수 있는 거래규칙을 생성한다. 이 때, 정확집합이라고 할 수 있는 경우는 입력변수들이 특정한 범주에 포함되어 있을 때, 입력 데이터의 결정속성이 모두 상승 또는 하락으로 분류되는 경우이다. 이외에 상승 또는 하락이 섞여 있는 거래규칙은 모두 러프집합이라고 볼 수 있는데, 본 연구에서는 상승 또는 하락 중 더 많이 포함되어 있는 방향으로 러프집합을 분류하였다.

본 연구에서는 조건속성으로 3개의 기술적지표를 변수선택하였는데 이는 Lee 등 (2010)의 연구에서 러프집합 분석을 통한 거래규칙 생성 시 3개의 지표를 조건속성으로 사용하는 것이 가장 적합함이 입증되었기 때문이다. Fugure 3.3은 일련의 과정에 대한 간단한 예시이다. 입력변수로 A, B, C라는 기술적지표가 선택되었을 때, 각 지표값은 상단과 하단처럼 다양한 방식으로 이산화된다. 이때, 이산화분위수와 이산화비율은 지표별로 다르게 적용되며, 이 값을 결정하는 것은 유전자알고리즘의 최적화변수이다. 범주형으로 표현된 조건속성과 과거데이터의 선물종가를 결정속성으로 하여 집합이 생성된다. 러프집합 분석은 이렇게 형성된 러프집합을 입력데이터를 기반으로 분류하여 거래규칙을 생성한다. 예를 들면, 지표A는 첫 번째 범주, 지표B는 두 번째 범주, 지표C는 첫 번째 범주에 해당하는 결정속성을 입력데이터를 기반으로 연산하였을 때 50번의 포함된 경우 중 30번의 경우가 상승, 20번의 경우가 하락을 나타내었다면 상승에 해당하는 거래규칙으로 결정된다. 이 거래규칙을 매매전략에 적용할 때는 A, B, C 지표가 해당 범주 안에 포함되게 되는 순간 선물종가의 상승을 예측하고 매수 시그널로 인식하게 된다.

본 연구에서는 하나의 거래규칙 집합을 만들기 위해 30번의 러프집합 분석을 반복하였다. 그 이유는 하나의 거래전략에 다양한 거래규칙들을 확보하여 거래규칙 집합의 예측력을 향상시키기 위함이다. 즉, Figure 3.2에서 볼 수 있듯이, 하나의 거래전략을 완성하기 위해 변수선택, 이산화, 러프집합 분석의 과정을 30회 반복하여 얻어낸 거래규칙의 집합이 하나의 거래전략으로 구성된다. 1개의 거래규칙 집합에 대한 유전자 알고리즘 최적화는 동시에 진행된다. 즉, 기술적지표 3개씩 30회에 해당하는 90개의 지표

에 대한 최적화변수가 하나의 유전자로 구성되어 유전자 알고리즘 최적화의 최적화변수로 사용된다.

생성된 거래규칙을 일정한 기준에 따라 거래에 활용할 유의미한 거래규칙을 선별한다. 선별 기준으로는 LHS 커버리지 (left-hand side coverage)와 RHS 정확도 (right-hand side accuracy)의 두 가지 기준을 사용한다. LHS 커버리지는 거래규칙의 좌변에 해당하는 조건속성이 러프집합 분석으로 분류된 전체 거래규칙중에 얼마만큼의 비중을 차지하는지를 의미한다. 즉, LHS 커버리지가 높은 거래규칙은 입력데이터 기준으로 빈도가 높은 거래규칙이기 때문에 거래전략 내에서 높은 영향력을 가지며, 이에 반해 LHS 커버리지가 낮은 거래규칙은 영향력이 적은 거래규칙이다. 본 연구에서는 LHS 커버리지의 하한을 제한하여 거래규칙을 선별하였는데, 일정 %이하의 커버리지를 가지는 거래규칙은 거래전략에 포함될 만큼 중요하지 않다고 판단하여 최종 거래규칙집합에서 제외시켰다. RHS 정확도는 거래규칙의 우변에 해당하는 결정속성이 가지는 방향성을 나타낸다. 본 연구에서 사용하는 결정속성은 1 (상승)과 -1 (하락)의 2가지 경우이기 때문에 RHS 정확도가 높을수록 해당 거래규칙의 입력데이터에서 1 또는 -1의 빈도가 높았음을 의미한다. RHS 정확도가 높은 거래규칙은 강한 방향성을 가지고 있는 것으로 해석되어 강한 예측력을 보이는 거래규칙이고, 이에 반해 RHS 정확도가 50%이면 해당 입력데이터에 1과 -1이 정확히 동등한 빈도로 측정되었으므로 방향성이 없는 것으로 볼 수 있다. 본 연구에서는 RHS 정확도의 하한에 대해 제한하여 낮은 예측력을 가진 거래규칙이 거래전략에 포함되지 않도록 선별하였다.

$$LHSCoverge = \frac{Number\ of\ LHS\ support}{Total\ number\ of\ LHS\ support} \tag{3.1}$$

$$RHSAccuracy = \frac{Number\ of\ RHS\ support}{Number\ of\ LHS\ support} \tag{3.2}$$

3.3. 거래 시뮬레이션을 통한 수익률 측정

본 연구에서 제안한 모델은 학습구간, 시뮬레이션 구간, 테스트 구간의 3구간으로 구성된다. 완성된 거래규칙 집합을 시뮬레이션 구간에 적용하여 거래 시뮬레이션을 수행하고 해당 거래규칙 집합의 수익률을 측정한다. 시뮬레이션 구간에서는 학습구간에서 생성한 거래규칙 집합을 적용하여 매매 시뮬레이션을 통해 수익률을 계산하고 해당 거래전략의 성과를 분석한다. 생성된 거래전략의 성과를 측정하는 척도로 시뮬레이션 구간의 수익률이 사용되며 이는 유전자 알고리즘 최적화을 위한 적합도 함수로도 사용된다. 학습구간에서 거래전략을 시뮬레이션하지 않고 새로운 가격정보구간을 사용하는 이유는, 학습구간에서 생성한 거래전략을 다시 학습구간에 적용할 때 과적합 (overfitting)이 발생하기 때문이다. 과적합이란, 데이터 기반의 인공지능 방법론에서 모델이 주어진 데이터에 과도하게 기반하여 모델을 다른 데이터에 적용할 때 적합한 성과를 보이지 못하는 것을 의미한다. 그러므로 거래전략 생성에 사용하는 데이터와 성과 측정에 사용하는 데이터를 다르게 설정하여 과적합을 줄이려고 노력하였다.

수익률은 측정하는 시뮬레이션은 거래전략에서 발생하는 매매시그널을 바탕으로 선물 포지션 매매를 통해 거래하였다. 특정 시점에서 선물가격에 대한 기술적지표가 거래규칙들의 조건속성 범주에 포함되면 해당 거래규칙의 예측값이 매매시그널로 발생된다. 즉, 기술적지표의 값들이 특정 거래규칙의 조건들을 모두 부합하는 상태일 때 결정속성을 통해 매매시그널을 발생한다. 시뮬레이션은 선물 1계약을 바탕으로 매수포지션, 매도포지션, 중립포지션의 총 3가지의 포지션을 취할 수 있다. 매수/매도포지션에서 반대포지션의 매매시그널이 발생하면 중립포지션으로 변경되며, 중립포지션에서 매매시그널이 발생하면 해당포지션으로 진입한다. 매매수익은 포지션 청산가격과 진입가격의 차이로 나타내었으며, 실제 매매와 유사하게 수익률을 측정하기 위해 거래 슬리피지 (slippage)를 적용하였다. 거래를 체결하기 위해서는 시장가보다 매수의 경우 매도호가, 매도의 경우 매수호가로 주문가를 지정해야 바로 주문이 체결되는데, 이때 호가와 시장가의 괴리로 발생하는 체결오차를 슬리피지라고 한다. 실제 매매에서 반드시

발생하는 손실부분에 해당하기 때문에 매매수익을 계산할 때 슬리피지를 차감하여 실제와 유사한 시뮬레이션 환경을 만들기 위해 노력하였다.

3.4. 유전자 알고리즘을 이용한 최적화

본 연구가 제안하는 거래모델은 주어진 구간에서 수익률을 극대화 할 수 있는 거래규칙 집합을 찾아내기 위하여 유전자 알고리즘을 이용하여 최적화 과정을 거친다. Figure 3.1에서 실제 완성된 거래규칙을 적용하여 매매에 활용하는 구간은 테스트 구간이지만, 본 모형은 시뮬레이션 구간의 시세정보에서 수익률을 극대화하도록 최적화된 거래규칙을 도출한다. 이렇게 접근하는 것에는 두 가지 이유가 있는데, 첫째는 실시간으로 거래규칙을 생성하는 것을 목표로 하고 현재시점까지의 데이터를 이용하여 거래규칙을 생성하기 위해서이다. 즉, 시뮬레이션 구간까지의 정보는 모형이 알고 있는 과거의 데이터이고, 테스트 구간에서의 데이터는 과거 데이터를 이용하여 생성한 거래규칙을 실제로 적용시키는 구간이기 때문이다. 둘째로 유전자 알고리즘을 이용하여 최적화한 거래규칙이 적어도 짧은 기간 동안은 여전히 효과적인 거래규칙일 것이라 생각되기 때문이다.

본 연구에서 사용되는 유전자 알고리즘에서 목적으로 하는 것은 시뮬레이션 구간에서의 수익률의 극대화이다. 최적화의 요인으로 사용되는 것은 크게 2가지인데, 러프집합 분석의 입력변수선택과 이산화과정의 이산화요인조절이다. 러프집합 분석은 하나의 거래규칙을 만들 때 기술적지표를 선택하여 입력변수로 사용하는데, 이 때의 변수선택을 염색체의 구성요소로 포함시켜 유전자 알고리즘이 선택하게 하였다. 따라서 수익률로 거래규칙을 최적화하게 되면, 해당 구간에서 가장 높은 수익률을 얻게 되는 기술적지표의 조합을 유전자알고리즘이 진화하며 능동적으로 선택하게 된다. 이산화과정에서 유전자에 포함시키는 이산화조절요인은 이산화분위수와 이산화비율이 있으며, 각 입력변수마다 다르게 적용된다. 동등비율로 이산화를 진행할 수도 있지만 각 기술적지표마다 특성이 다르기 때문에 유전자 알고리즘이 진화함에 따라 각 지표에 적합한 이산화분위수와 이산화비율이 선택되게 된다.

4. 실증분석

본 연구에 사용된 데이터는 코스콤에서 제공하는 프로그램 CHECK Expert로부터 구하였으며 국내 KOSPI200 주가지수 선물의 시가, 고가, 저가, 종가, 거래량에 대한 30분 간격의 시세데이터를 사용하였다 (Kim, 2010). 본 연구가 제안하는 모델은 거래규칙 생성구간, 시뮬레이션 구간, 테스트 구간의 총 3가지 구간으로 데이터 구간을 지정하여 사용하는데, 각각 2개월, 1개월, 1개월로 지정하여 사용하였다.

입력변수로 사용되는 34개의 기술적지표는 Lee와 Ahn (2010)의 최근 연구에서 선물시장에 적합하다고 알려진 기술적지표들을 참고하여 구성하였고, Table 4.1에서 상세히 기술하였다. 거래규칙 선별단계에서 사용하는 선별기준으로는 LHS Coverage 1%이상, RHS Accuracy 55% 이상을 사용하였다.

완성된 거래규칙 집합을 이용한 거래 시뮬레이션은 하나의 선물계약 포지션을 이용하여 매수 포지션 혹은 매도 포지션으로 거래하는 전략을 시뮬레이션 하였다. 시뮬레이션을 실제 거래와 유사하게 하기 위한 슬리피지는 매 주문시마다 1틱, 즉 0.05pt를 손익에서 차감하였다.

유전자 알고리즘 최적화를 수행하기 위한 유전자는 변수선택, 이산화분위수, 이산화비율에 대한 정보를 가지고 있다. 변수선택을 위한 값은 각각 1∼34 사이의 3개의 정수가 필요하며 이는 러프집합 분석을 위해 선택된 기술적지표의 번호를 의미한다. 이산화분위수는 2∼5 사이의 3개의 정수가 필요하여 각 기술적지표를 몇 분위로 이산화할 것인지를 의미한다. 마지막으로 이산화비율에 대한 정보는 최대 12개 ((5분위시 4개) * (3개의 기술적 지표))의 실수가 필요하며, 각각 이산화 구간 간의 크기의 비율을 의미한다. 이산화비율은 상대적인 비율이기 때문에 0∼1 사이의 실수값으로 결정된다.

러프집합 분석과 유전자 알고리즘 최적화를 포함한 본 연구를 수행하기 위한 프로그램은 C++ 언어를 기반으로 직접 제작하여 사용하였다. 이때, 러프집합 분석은 Uppsala 대학의 ROSETTA 라이브러

리 (http://www.lcb.uu.se/tools/rosetta)를 사용하였고, 유전자 알고리즘은 MIT의 GALib 라이브러리 (http://lancet.mit.edu/ga/)를 사용하여 구현하였다.

```
Price_Oscillator([...,... ))  AND Volatility([*,... ))  AND VROC([...,.. ))  = -1 and 0
Price_Oscillator([..., *)) AND Volatility([*,... ))  AND VROC([*,... ))     = 1 and 0
Price_Oscillator([..., *)) AND Volatility([*,... ))  AND VROC([...,..))     = 1 and 0
Sonar_signal([*,... ))  AND MFI([..., *)) AND MACD([*,... ))                = -1 and 0
Sonar_signal([*,... ))  AND MFI([..., *)) AND MACD([..., *))                = -1 and 0
```

Figure 4.1 Example of trading strategy rules

Figure 4.1은 거래모형에서 최종적으로 완성되는 거래규칙의 예시이다. 거래규칙의 조건은 사용된 기술적지표의 종류와 이산화된 구간으로 표현된다. 거래규칙의 예측값은 1 또는 -1과 0으로 구성되는데, 현재의 기술적지표 값들이 좌측 조건의 범위에 포함되는 이는 1 (매수시그널) 혹은 -1 (매도시그널)로 판단된다. 0은 해당 조건에 포함되지 않는 경우 포지션을 유지함을 의미한다. 본 연구에서 제안한 알고리즘 트레이딩 매매 전략을 바탕으로 과거 주가지수 선물시장의 데이터를 바탕으로 거래성과를 측정하였다. 테스트 기간은 2009년 1월부터 2012년 12월까지 총 4년간의 기간을 사용하였으며, 슬라이드 윈도우 (slide window) 방식으로 1개월 간격으로 이동하여 총 48번의 실험을 수행하였다.

Table 4.2은 러프집합 이론과 유전자 알고리즘을 이용하여 산출된 학습구간에서의 거래규칙 수익률이다. 유전자 알고리즘이 학습구간에서 수익률을 극대화하도록 최적화하고 최고의 수익률을 보이는 거래규칙을 찾아낸다. 매달 측정된 누적수익률은 KOSPI200선물의 수익포인트로 나타내었다. 2009년부터 2012년의 4년간 매년 163.75pt∼230.4pt의 매우 높은 연간 수익포인트를 기록한 것으로 보아, 학습구간에서 유전자 알고리즘을 통한 최적화가 올바르게 진행된 것을 확인 할 수 있었다. 특히 48번의 시뮬레이션 중 2번을 제외하고는 양의 수익포인트를 나타낸 것으로 보아, 많은 거래규칙들 중 현재 시장상황에서 수익을 기대할 수 있는 거래규칙이 높은 확률로 존재할 수 있음을 알 수 있다. Table 4.3은 학습구간에서 최적해로 선택된 거래규칙들을 바로 다음 1개월 동안 적용시켜 실제 매매전략의 수익률을 측정하였다. 2009년은 대세상승장으로, 매매전략이 효과적으로 적용되어 52.3pt의 높은 수익을 보였으며, 2010년과 2012년에도 유의미한 수익을 올리는 것으로 확인되었다. 2011년의 경우, 8월에 미국 신용등급 강등으로 인한 글로벌 시장 및 국내시장의 급락장이 형성되었는데, 기술적 분석의 고유적인 특성상 시장변화에 후행하는 성질에 때문에 적절히 대응하지 못한 것으로 보이며 향후 연구를 통해 개선해야할 부분이다.

Table 4.4는 제안한 매매전략의 성과를 측정하기 위해 샤프지수 (shape ratio)를 측정하였다. 샤프지수는 투자전략의 성과를 측정하기 위해 고안된 지수로, 투자를 통해 얻게 된 수익률에서 무위험 수익률 (risk-free rate)를 차감한 초과수익률을 수익률의 변동성으로 나누어 전략의 위험대비 수익률을 측정하는 지수이다. 샤프지수의 성과를 상대적으로 비교하기 위하여 벤치마크지수로 매수 후 보유 전략 (buy and hold) 전략을 사용하였다. 매수 후 보유 전략은 투자의 성과를 비교하기 위하여 흔히 사용되는 벤치마크지수로, 매매의 시작시점에 매수를 한 뒤 마지막시점에서 매도하여 그 투자수익를 측정하는 전략이다. 샤프지수를 위한 변동성 측정은 제안 전략이 1개월 간격으로 거래 규칙을 변경하는 바, 한 달 동안의 수익률을 측정하고 1년간의 수익률들의 표준편차를 사용하였다. Table 4.4에서 볼 수 있듯이, 매수 후 보유 전략보다 제안한 매매전략이 개선된 샤프지수를 보임을 알 수 있었다. 단, 2009년의 경우에는 KOSPI200 선물시장이 급격한 상승장을 보이어 매수 후 보유 전략의 초과수익률이 매우 높아 상대적으로 낮은 샤프지수를 보였다. 이외의 구간에서는 상대적으로 높은 샤프지수를 보여, 제안한 매매전략이 시장의 단순 변화보다 효율적인 투자전략임이 확인되었다.

Table 4.1 Technical indicators for input variable

MAO (MA Oscillator)	$MAO_t = MA_t(n) - MA_t(m), n < m$	EOM	$EOM = \frac{\frac{H_t + L_t}{2} - \frac{H_{t-1} - L_{t-1}}{2}}{\frac{Vol_t}{H_t - L_t}}$
TRIX	$TRIX_t(n) = \frac{EMA_t^3(C,n) - EMA_{t-1}^3(C,n)}{EMA_{t-1}^3(C,n)}$	OBV (On Balance Volume)	$OBV_t = \begin{cases} OBV_{t-1} + V_t, C_t > C_{t-1} \\ OBV_{t-1} - V_t, C_t < C_{t-1} \\ OBV_{t-1}, C_t = C_{t-1} \end{cases}$
MI (Mass Index)	$MI_t(n) = \sum_{i=1}^{n} \frac{EMA_{t-n+i}(r,9)}{EMA_{t-n+i}^2(r,9)}$	William's	$William' = \frac{H_{r,\max(n)} - C_t}{H_{t,\max(n)} - L_{t,\min(n)}} \times 100$
Momentum	$Momentum_t(n) = \frac{C_t - L_{t-n}}{H_{t-n} - L_{t-n}} \times 100$	CO (Chaikin's Oscillator)	$CO_t(m,n) = EMA_t(AD, m) - EMA_t(AD, n)$
Stochastic	$Stochastic_t(n) = \frac{C_t - L_{t-n}}{H_{t-n} - L_{t-n}} \times 100$	Disparity	$Disparity = \frac{C_t}{MA_t(n)} \times 100$
SMI (Stochastic Momentum Index)	$SMI_t(n) = 100 \times \left(\frac{2C_t}{\left(H_{t,\max(n)} - L_{t,\min(n)}\right)} - 1 \right)$	SI (Swing Index)	$SI_t = \frac{50 \times K_t}{M_t} \times \left(\frac{\left(C_t - C_{t-1}\right) + 0.5(C_t - O_t) + 0.25(C_{t-1} - O_{t-1})}{R} \right)$ $K = MAX(H_t - C_{t-1}, L_t - C_{t-1})$ $R = MAX(H_t - C_{t-1}, L_t - C_{t-1}, H_t - L_t)$
VROC	$VROC_t(n) = \left(\frac{V_t}{V_{t-n}} - 1 \right) \times 100$	Zscore	$Zscore_t = \frac{C_t - MA_t(n)}{STDEV_t(n)}$
ROC	$ROC_t(n) = \left(\frac{C_t}{C_{t-n}} - 1 \right) \times 100$	TP (Typical Price)	$TP_t = \frac{C_t + L_t + H_t}{3}$
RSI (Relative Strength Index)	$RSI_t(n) = 100 - \frac{100}{1 + RS_t(n)}$ *where* $RS_t(n) = \sum_{i=1}^{n} U_{t-n+i} \sum_{i=1}^{n} D_{t-n+i}$ *where* $U_t = \begin{cases} C_t - C_{t-1}, C_t \geq C_{t-1}, \\ 0, otherwise \end{cases}$ $U_t = \begin{cases} C_t - C_{t-1}, C_t \geq C_{t-1}, \\ 0, otherwise \end{cases}$	PO (Price Oscillator)	$PO_t(m, n) = \frac{SMA_t(m) - SMA_t(n)}{SMA_t(m)}$
DMI (Directional Movement Indicators)	$+DM_t = \begin{cases} H_t - H_{t-1} \ for \ H_t - H_{t-1} > 0, \\ H_t - H_{t-1} > L_{t-1} - L_t \\ 0, otherwise \end{cases}$ $-DM_t = \begin{cases} L_t - L_{t-1} \ for \ L_t - L_{t-1} > 0, \\ L_t - L_{t-1} > L_{t-1} - L_t \\ 0, otherwise \end{cases}$	MFI (Money Flow Index)	$MFI_t(n) = 100 - \frac{100}{1 + MF_t}$ *where* $MF_t = \frac{PMF_t}{NMF_t}$ *where* $PMF_t = \begin{cases} PMF_t + C_t - C_{t-1}, C_t \geq C_{t-1}, \\ 0, otherwise \end{cases}$ $NMF_t = \begin{cases} NMF_t + C_{t-1} - C_t, C_t \leq C_{t-1}, \\ 0, otherwise \end{cases}$
ADX (Average Directional Index)	$\begin{cases} ADX_t(n) = \overline{DX}_t(n) \ where \\ DX_t = \frac{\lvert(+DI_t) - (-DI_t)\rvert}{(+DI_t) + (-DI_t)} \times 100 \\ +DI_t = +DM_t / TR_t \\ -DI_t = -DM_t / TR_t \\ TR = MAX(H_t - L_t, \lvert C_{t-1} - C_t \rvert, \lvert C_{t-1} - L_t \rvert) \end{cases}$	VO (Volume Oscillator)	$VO_t(n, m) = \frac{\bar{V}_t(m) - \bar{V}_t(n)}{\bar{V}_t(n)} \times 100$
ADXR (Average Directional Index Rating)	$ADXR = A\bar{D}X(n)$	DP (Detrended Price)	$DP_t = C_t - MA_{t-(n/2+1)}(n)$
CCI (Commodity Channel Index)	$\begin{cases} CCI_t(n) = \frac{M_t - \bar{M}_t(n)}{d_t(n) \times 0.015} \\ where \ M_t = (H_t + L_t + C_t)/3 \text{ and} \\ d_t(n) = \frac{1}{n} \sum_{i=1}^{n} \lvert M_{t-n+i} - \bar{M}_t(n) \rvert \end{cases}$	NCO (Net Change Oscillator)	$NCO_r(n) = C_t - C_{t-n}$
MACD	$MACD_t(m,n) = EMA_t(C,m) - EMA_t(C,n)$	EMA (Exponential Moving Average)	$EMA_t(y, n) = \frac{\sum_{i=1}^{n} a^{n-i} y_{t-n+i}}{\sum_{i=1}^{n} w_i}$ *where* $a = exponential \ factor \ (0 < a < 1)$

Table 4.2 Monthly return on training period (Pt)

	2009	2010	2011	2012
Jan.	32.65	23.2	24.85	5.65
Feb.	25.3	6.1	6.9	29.45
Mar.	3.25	11.1	-6.85	20.5
Apr.	34.1	17.1	37.1	7.85
May.	27.65	15	22.15	9.8
Jun.	11.1	10.2	8.7	-4.95
Jul.	13.2	19.6	11.5	18.4
Aug.	28.65	20.55	13.8	21.35
Sep.	12.9	6.95	12.55	13.9
Oct.	21.85	20.05	16.1	23.45
Nov.	7.45	9.3	46.6	2.5
Dec.	12.3	20.5	23.5	15.85
Cumulative	**230.4**	**179.65**	**216.9**	**163.75**

Table 4.3 Monthly return on testing period (Pt)

	2009	2010	2011	2012
Jan.	-2.65	-12.35	-7.55	19.4
Feb.	-8.9	-6.95	-19.5	7.3
Mar.	11.55	5.9	29.45	-0.8
Apr.	6.2	5	12.75	-0.75
May.	7.7	-8.75	-2.9	-26.3
Jun.	2.2	0.25	-10.35	6.8
Jul.	21.65	5.7	0.45	1.2
Aug.	4.6	-9	-30.35	1.85
Sep.	15.05	16	-0.9	14.15
Oct.	-21.6	1.3	19.9	-8.15
Nov.	0.5	0.05	-6.4	-1.45
Dec.	16	15.25	-13.2	10.25
Cumulative	**52.3**	**12.4**	**-29.6**	**23.5**

Table 4.4 Sharpe ratio on testing period

	Purposed Model	Buy and Hold
2009	1.220	1.605
2010	1.567	1.118
2011	-0.584	-0.677
2012	0.704	0.555

5. 결론

국내 KOSP200 선물시장의 거래규모와 변동성은 빠른 속도로 증가해왔다. 파생상품인 선물시장의 특성 상 기존 주식시장의 투자 분석 방법이외에 선물시장만의 특성을 반영하는 투자전략에 대한 연구의 필요성이 대두되고 있다. 특히 선물시장은 높은 레버리지 효과로 인해 투자의 위험이 매우 높아 투자자의 주관적 판단을 통한 매매기법보다 과거 데이터를 기반으로 합리적인 매매전략을 구사하는 알고리즘 트레이딩 기법에 대한 중요성이 강조되고 있다. 본 연구에서 제안한 인공지능 방법론을 이용한 매매전략은, 거래규칙을 생성하는 러프집합 이론과 유전자 알고리즘을 이용한 최적화의 2가지 인공지능 방법론을 융합하여 거래규칙 생성 모형을 제안한 점에서 시사하는 바가 크다. 유전자 알고리즘을 이용하여 해당 구간에서 가장 높은 수익률을 보이는 거래규칙을 생성하여 러프집합을 이용한 거래규칙의 효과를 극대화한 것이 본 연구의 특징이다. 실증 분석 결과, 제안한 매매전략이 매수 후 보유 전략보다 높은 샤프지수를 나타내어 실효성있는 투자전략임도 입증되었다.

본 연구의 한계점으로는 기술적지표의 특성상 시장의 급격한 외부충격으로 급락장이 발생했을 때 급

격한 변화가 선물가격정보에 반영된 이후에 기술적지표에 반영되기 때문에 후행적 성질을 가지고 있다는 점이다. 이 문제점을 개선하기 위해 본 연구에서 사용한 30분 간격 데이터보다 짧은 간격인 1분 간격 데이터, 틱 (tick) 데이터 등을 사용하여 기술적지표를 연산한다면 급락장에 보다 긴밀하게 대응할 수 있을 것이라 생각된다.

References

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 Korean Data & Information Science Society*, **20**, 1049-1060.

Chang, P. C. and Liu, C. H. (2008). A TSK type fuzzy rule based system for stock price prediction. *Expert Systems with Applications*, **34**, 135-144

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.

Dai, J. H. and Li, Y. X. (2002). Study on discretization based on rough set theory, *Preceedings of the First International Conference On Machine Learning and Cybernetics*, 4-5.

Dong, M. and Zhou, X. S. (2002). Exploring the fuzzy nature of technical patterns of U.S stock market. *ICONIP'02-SEAL'02-FSKD'02*, Singapore, 18-22.

Fazel Zarzndi, M. H., Rezaee, B., Turksen, I. B. and Neshat, E. (2009). A type-2 fuzzy rule-based expert system model for stock price analysis. *Expert Systems with Applications,* **36**, 139-154.

Fong, S., Tai, J. and Si, Y. W. (2011). Trend following algorithms for technical trading in stock market. *Journal of Emerging Technologies in Web Intelligence*, **3**, 136-145.

Kang, Y. J. and Oh, K. J. (2013). Using rough set to develop a volatility reverting strategy in options market. *Journal of Korean Data & Information Science Society*, **24**, 135-150.

Kim, H, H. and Oh, K. J. (2012). Using rough set to develop the optimization strategy of evolving time-division trading in the futures market. *Journal of Korean Data & Information Science Society*, **23**, 881-893.

Kim, K. J. and Han, I. G. (2000). Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert Systems with Applications*, **19**, 125-132.

Kim, S. and Ahn, H. (2010). Development of an intelligent trading system using support vector machines and genetic algorithms. *Journal of Intelligence and Information Systems*, **16**, 71-92.

Kuo, R. J., Chen, C. H. and Hwang, Y. C. (2001). An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network. *Fuzzy Sets and Systems*, **118**, 21-45.

Lee, S. J., Ahn, J. J., Oh, K. J. and Kim, T. Y. (2010). Using rough set to support investment strategies of real-time trading in futures market. *Applied Intelligence*, **32**, 364-337.

Lo, A. W., Mamaysky, H. M. and Wang, J. (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *Journal of Finance*, **55**, 1705-1770.

Marc, B. (2004). Khiops: A statistical discretization method of continuous attributes. *Machine Learning*, **55**, 53-69.

Pawlak, Z. (1997). Rough set approach to knowledge-based decision support. *European Journal of Operational Research*, **99**, 48-57.

Journal of the Korean Data & Information Science Society
2014, **25**(2), 281–292

http://dx.doi.org/10.7465/jkdi.2014.25.2.281
한국데이터정보과학회지

Using genetic algorithm to optimize rough set strategy in KOSPI200 futures market

Seung Hwan Chung[1] · Kyong Joo Oh[2]

[12]Department of Information and Industrial Engineering, Yonsei University

Received 2 December 2013, revised 26 December 2013, accepted 27 January 2014

Abstract

As the importance of algorithm trading is getting stronger, researches for artificial intelligence (AI) based trading strategy is also being more important. However, there are not enough studies about using more than two AI methodologies in one trading system. The main aim of this study is development of algorithm trading strategy based on the rough set theory that is one of rule-based AI methodologies. Especially, this study used genetic algorithm for optimizing profit of rough set based strategy rule. The most important contribution of this study is proposing efficient convergence of two different AI methodology in algorithm trading system. Target of purposed trading system is KOPSI200 futures market. In empirical study, we prove that purposed trading system earns significant profit from 2009 to 2012. Moreover, our system is evaluated higher shape ratio than buy-and-hold strategy.

Keywords: Algorithm trading, futures market, genetic algorithm, rough set.

[1] Graduate student, Department of Information and Industrial Engineering, Yonsei University, Seoul 120-749, Korea.
[2] Corresponding author: Associate professor, Department of Information and Industrial Engineering, Yonsei University, Seoul 120-749, Korea. E-mail: johanoh@yonsei.ac.kr