

유전 알고리즘에 의한 종목 추천에 관한 연구

구규림, 박정우, 진민재, 최준수*

국민대학교 컴퓨터공학부

e-mail : gyulim, jungwooc, mj_jeon, jschoi@kookmin.ac.kr

A Study on Genetic Algorithm for Recommending Stocks

Gyulim Gu, Jungwoo Park, MinJae Jeon, Joonsoo Choi

Dept. of Computer Science, Kookmin University

요 약

유전 알고리즘 (Genetic Algorithm)은 기존의 알고리즘 개발방법을 통하여 해결하기 어려운 최적화 등의 문제를 해결하기 위한 자연계의 진화과정을 모방한 방법이다. 본 연구에서는 유전 알고리즘을 이용하여 KOSPI 200에서 거래되고 있는 증권의 매수/매도 종목을 추천하는 방법을 제시한다. 이를 위하여 기술적 분석 (Technical Analysis) 방법 중에서 Slow Stochastic 지표와 MACD 지표를 이용하여, 매일매일 두 지표가 나타내는 매매 신호를 기반으로 해당하는 각각의 종목에 대해 최근 가장 좋은 수익률을 나타내는 매수/매도 종목을 추천하는 방법을 구현한다.

1. 서론

주식 투자에 있어 기술적 분석은 과거의 주가와 거래량을 이용해 주식의 추세를 분석하여 미래의 주가를 예측하는 방법이다[1]. 그러나 기술적 분석을 통한 방법은 완벽하게 믿을 수 있는 것은 아니다. 가령 새로운 기술을 개발했거나 국가적 위기 상황과 같은 요건들은 반영할 수 없다. 그럼에도 불구하고 기술적 분석은 위험성이 높은 주식 투자에서 안정적인 방법으로 간주되며 종목의 추세를 반영하기 때문에 장기 투자에서는 상당한 효과를 발휘한다.

기술적 분석은 지표를 계산함에 있어서 임의적인 기간이 필요하다. 예를 들어 MACD 지표를 구하기 위해서는 단기 이동평균, 장기 이동평균 등의 기간을 설정해야한다. 따라서 각각의 종목마다 주가의 흐름은 다양하기 때문에 같은 기간의 지표를 이용해 모든 종목들에 대해 기술적 분석을 하는 것은 효율적인 방법이 아니다. 따라서 종목마다 종목에 맞는 지표를 그리고 이를 이용해 기술적 분석을 하여 투자를 한다면 효과적으로 수익을 얻을 수 있을 것이다.

본 연구에서는 유전 알고리즘[2-4]을 통해 각각의 종목에 알맞은 지표를 구하고 이를 통하여 기술적 분석을 기반으로 매수/매도 종목을 추천하는 시스템을 구현하고자 한다.

2. 유전 알고리즘과 기술적 분석

본 연구에서는 기술적 분석인 Slow Stochastic 지표와 MACD (Moving Average Convergence-Divergence) 지표[1]를 사용한다. 두 지표를 이용해 유전 알고리즘을 구성하는 방법과 이를 통해 종목을 추천하는 방향을 제시한다.

2.1 기술적 분석

유전 알고리즘의 유전자를 구성하는데 있어서 기술적 분석에서 많이 사용하고 있는 지표인 Slow Stochastic과 MACD를 사용한다. 두 지표는 실무에서 많이 사용되며 신뢰도가 높다. 두 지표를 구하기 위한 방법은 아래와 같다.

MACD는 수식(1)에서와 같이 단기 EMA (Exponential Moving Average, 지수이동평균)와 장기 EMA의 차이이다. 단기이동평균으로는 12일을, 장기이동평균으로는 26일을 많이 사용한다. MACD Signal 은 k (주로 9일) 일간의 지수이동평균을 나타낸다. 또한 MACD Oscillator 는 MACD와 MACD Signal 과의 차이를 나타낸다.

$$\begin{aligned}
 \text{MACD Line} &= \text{단기 EMA} - \text{장기 EMA} \\
 \text{MACD Signal} &= k \text{기간의 EMA} \quad \dots(1) \\
 \text{MACD Oscillator} &= \text{MACD} - \text{MACD Signal}
 \end{aligned}$$

* 본 연구는 국민대학교의 교내연구비 지원 사업으로 수행된 연구결과임

Stochastic 지표는 주가 흐름으로 만들어지는 파동의 성격을 지표화한 것으로 Fast, Slow Stochastic 의 두 가지 종류가 있다. Slow Stochastic 은 아래 수식(2)에서와 같이 Fast Stochastic 을 평균이동한 지표이다.

$$\text{Fast\%}K = \frac{\text{현재 증가} - N\text{기간 저가}}{N\text{기간 고가} - N\text{기간 저가}} \times 100$$

$$\text{Fast\%}D = \text{Fast\%}K \text{의 } M\text{기간 이동평균} \quad \dots(2)$$

$$\text{Slow\%}K = \text{Fast\%}D$$

$$\text{Slow\%}D = \text{Slow\%}K \text{의 } n\text{기간 이동평균}$$

본 연구에서는 위의 두 지표를 계산하는데 사용되는 기간을 기존에 많이 사용하는 기간으로 고정하지 않고, 각 종목에 가장 적합한 날짜를 직접 구하여 계산하고자 한다. 따라서, MACD 지표에서는 단기, 장기, k의 3개의 날짜가 필요하며, Slow Stochastic 지표를 구하기 위해서는 N, M, n의 3개의 날짜가 필요하다. 여기에서 정의된 6개의 날짜 변수를 이용하여 유전자를 구성하고, 유전 알고리즘을 통해 종목에 가장 알맞은 유전자, 즉 6개의 날짜를 구한다.

2.2 유전자 구조

유전 알고리즘을 구현하기 위해서는 먼저 최적화 하고자 하는 문제의 해답이 될 수 있는 후보해를 코딩하고 있는 유전자의 구조를 정하여야 한다. MACD와 Slow Stochastic 의 최적화를 위해서 두 지표의 각각의 날짜를 변수로 두고 그 범위를 아래와 같이 지정한다.

(표 1) 유전자 범위

x1(N)	1	<=	x1	<=	127
x2(M)	x1	<=	x2	<=	256
x3(n)	1	<=	x3	<=	256
x4(단기)	1	<=	x4	<=	256
x5(장기)	1	<=	x5	<=	256
x6(k)	1	<=	x6	<=	256

위 여섯개의 변수를 각각 이진화하고 8개의 비트로 표현하면, 아래와 같은 48비트의 유전자를 만든다. 이 유전자를 이용하여 날짜 변수로 한 최적의 Slow Stochastic와 MACD 지표를 만든다.

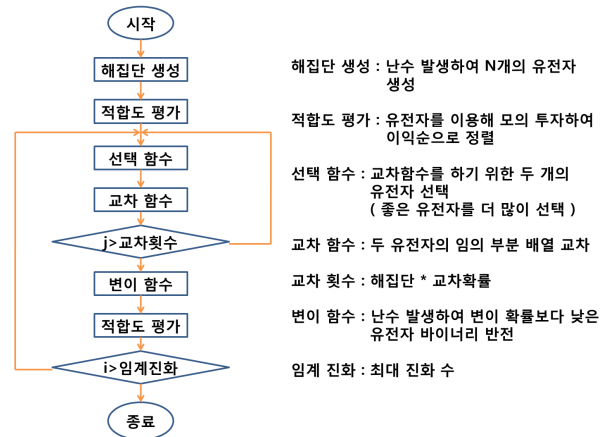
(표 2) 유전자 예시

00001100	00011010	00001001	00000011	00000100	00001000
x1	x2	x3	x4	x5	x6
(x1, x2, x3, x4, x5, x6) = 12, 26, 9, 3, 4, 8					

2.3 유전 알고리즘 흐름도

일반적인 유전 알고리즘은 아래와 같은 방법으로 진행된다. (그림 1 참조) [2].

- (1) 최초의 난수를 발생시켜 P개의 유전자인 해의 집단을 만든다.
- (2) 해의 집단의 적합도를 평가한다. 적합도란, 각각의 유전자를 이용해 Slow Stochastic와 MACD를 구하고 두 지표가 제공하는 매매시점을 이용해 특정 기간 동안 모의 투자를 하여 수익률을 계산한다.
- (3) 품질 비례 선택과, 엘리트 보전 전략을 이용해 교배할 두 개체들을 선택한다. 개체를 선택하여 교배하는 연산은 (개체 집단의 크기 * 교배 확률)의 수만큼 한다.
- (4) 해의 집단의 각각의 개체에 대해 난수를 발생시켜 변이 확률보다 적게 나올 경우 변이한다.
- (5) 교배와 변이 연산이 끝나면 다시 해의 집단의 적합도를 평가한다.
- (6) 종료 조건을 만족하면 다음단계로, 아니면 진화를 반복한다.
- (7) 진화된 해의 집단에서 가장 큰 수익률을 주는 해를 찾는다.



(그림 1) 유전 알고리즘 흐름도

2.4 적합도 평가

위 유전 알고리즘을 수행하면서 해집단의 개체들을 이용하여 Slow Stochastic와 MACD 지표를 구한 다음, 두 지표가 가리키는 매매 시점을 이용하여 최근 4년 동안 모의 투자하여 수익을 계산한다. 모든 개체들의 수익이 계산되면 초기 자산 대비 증가한 수익률을 구하고 이를 내림차순 정렬하여 0번째 개체가 가장 좋은 수익률을 가지도록 한다.

모의 투자는 증가에 대해서만 매매가 가능하며 시작 날의 증가를 초기 자산으로 가정한다. 우선적으로 Slow Stochastic 지표가 가리키는 매수/매도 신호로 모의 투자를 수행한다. 그러나, 주식을 보유 중에 MACD 지표가 매도 신호를 나타내거나 주식이 없는 상태에서 MACD 지표가 매수 신호를 나타내면 MACD 신호에 따라 투자하여 Slow Stochastic이 놓치는 시점을 보완한다.

증권사에 지불해야 하는 매수, 매도의 수수료와 매도 시 세금을 반영하여 모의 투자에서 이익을 계산하는데 실제 거래와 같도록 아래와 같은 방법을 적용한다.

```

/*
지표 신호
BUY:매수, SELL:매도 HOLD:보유, READY:대기
날짜 마다 지표 신호 저장
*/
int slowStoch[], macd[];
double profit = 0; // 수익
double close[]; // 종가
final double B = 0.00015; // 매수 수수료
final double S = 0.00015; // 매도 수수료
final double T = 0.003; // 세율
final double F = 0.035/365; // 순수무위험이자율

for (int i = 0; i < end; i++ ) {
    if ( slowStoch[i] == BUY )
        profit -= close[i] + ( close[i] * B );
    else if ( slowStoch[i] == SELL )
        profit += close[i] - close[i]*S - close[i] * T;
    else if ( macd[i] == BUY )
        profit -= close[i] + (close[i] * B );
    else if ( macd[i] == SELL )
        profit += close[i] - close[i] * S - close[i] * T;
    else
        profit = profit + ( profit * F );
}
    
```

(그림 2) 모의투자 수익계산 코드

2.5 선택 함수

수익률이 높은 개체가 선택 시 더 많이 선택 되도록 하여 가장 좋은 엘리트 유전자는 보존 되도록 한다. 따라서 최대 이익의 적합도가 최소 이익의 적합도의 K배가 되도록 수익률에 대한 적합도를 조정한다.

최대 이익을 pB , 최소 이익을 pW , 적합도를 F 라 하였을 때, 적합도 조정 공식은 아래 수식(3)과 같이 설정한다.

$$F[i] = p[i] - pW + \frac{(pB - pW)}{K - 1} \quad \dots(3)$$

위 수식(3)을 이용해 각각의 개체의 적합도를 조정 한 뒤 아래와 같은 함수를 이용하여 교배 연산에 사용할 두 유전자를 고른다.

```

//sum_of_F = F[]의 총합
point = rand(1, sum_of_F);

for ( int i = 0; i < F.length; i++ ) {
    sum += F[i];
    if ( point <= sum )
        return i;
}
    
```

(그림 3) 유전자 선택 함수

2.6 교배 함수

먼저 선택 함수를 통해 48비트의 두 유전자를 고른다. 다음으로 0 ~ 48 사이의 임의의 두 지점을 발생시킨 뒤 두 지점 사이의 비트를 바꾼다. 교배 과정은 다음과 같다.

초기 두 유전자	유전자 교배
X : 00001111,	X : 00110011,
Y : 11110000	Y : 11001100
(X : 15, Y : 240)	(X : 51, Y : 204)
※ 임의의 두 지점 to = 2, from = 5	

(그림 4) 유전자 교배과정

2.7 변이 함수

변이 과정에서는 난수를 발생 시켜 설정된 변이 확률 보다 작게 나왔을 경우 유전자 구조를 0 <-> 1로 변형한다. 변이 과정의 예는 아래와 같다.

초기 유전자	유전자 변이
X : 00001111	X : 11110000

(그림 5) 유전자 변이과정의 예

2.8 유전 알고리즘 실행 결과

위와 같은 방법으로 구현한 유전 알고리즘을 삼성전자 종목에 대해 2009년부터 2012년 까지 해집단의 크기 300, 진화 수 300, 교배확률 0.9, 변이확률 0.05로 유전 알고리즘을 수행한 결과는 아래와 같다.

Learning : 2006.01.02 ~ 2012.09.14
Best Gene
Short-MA : 6, Long-MA : 92, MACD-MA : 134
Fast %K : 145, Slow %K : 186, Slow %D : 222
Paper Trade 2009.02.17 ~ 2012.09.14
Start : 490,500.0원 -> End : 1,403,258.073 원
Profit : 912,758.073, Last Close : 1,336,000원
Rate of Profit : 186.087 %

(그림 6) 삼성전자 종목에 적용한 유전 알고리즘 결과

3. 종목 추천

KOSPI200 종목 등과 같은 추천하고자 하는 종목을 대상으로 매월마다 각 종목들에 대해 유전 알고리즘을 여러 번 수행시켜 그 중에서 가장 좋은 수익률을 나타내는 유전자를 데이터베이스에 저장한다. 매일 주식 시장이 마감하면 각 종목에 대해 저장한 유전자를 가지고 오늘까지의 Slow Stochastic과 MACD 지표를 구하여 다음날의 매매 신호를 파악한다. 이 중에서 매수와 매도신호가 발생한 종목들만 추출하여 매수를 추천하는 종목과 매도를 추천하는 종목을 제공한다.

하게 하기위해 종목의 수를 줄이고 매일 유전 알고리즘을 통해 가장 좋은 유전자를 갱신하며 매시간 마다 또는 매분마다 기술적 분석하여 매수 매도 종목 추천 목록을 구한다면 보다 유연하게 시장의 흐름에 대처하여 단기 투자에도 도움을 주는 투자전략이 될 수 있을 것이다.

앞으로 MACD 와 Slow Stochastic 이외의 다양한 지표를 적용하여 보다 정확한 전략을 수립하는 방법에 관하여 연구할 필요가 있으며, 유전 알고리즘 이외의 유전 프로그래밍 혹은 인공지능망 (Neural Network) 등을 적용한 주식 추천 알고리즘[3, 5] 에 관해서도 연구할 필요가 있다.

참고문헌

- [1] Edwards, R.D., and Magee, J., *Technical Analysis of Stock Trend*, American Management Association, 1998.
- [2] Goldberg, D.E., *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, 1989.
- [3] Anthony, B., and Michael, O., *Biologically Inspired Algorithms for Financial Modelling*, Springer, 2005.
- [4] 조희연, 김영민, 유전자 알고리즘을 이용한 주식투자 수익률 향상에 관한 연구, 한국정보시스템학회, 제12권 2호, pp 1-20, 2003.
- [5] Franklin A., and Risto K., Using Genetic Algorithms to Find Technical Trading Rules, *Journal of Financial Economics*, Vol 51, pp 245-271, 1999.

매수를 추천하는 종목

종목	수익률	기간
(주)케이티	35.108%	2008.03.04 ~ 2012.09.05
(주)케이티앤지	53.864%	2008.03.04 ~ 2012.09.05

매도를 추천하는 종목

종목	수익률	기간
(주)두산	89.583%	2008.02.29 ~ 2012.09.05
(주)LG상사	118.996%	2008.02.29 ~ 2012.09.05
일성신약(주)	65.243%	2008.03.03 ~ 2012.09.05
한일이화(주)	474.433%	2008.03.03 ~ 2012.09.05
현대산업개발(주)	37.5%	2008.03.03 ~ 2012.09.05

(그림 7) 2012년 9월 6일 매수/매도 종목

위 그림 7은 2012년 9월 5일 이전의 약 4년 5개월 동안의 가격에 대하여 위에서 제안한 유전 알고리즘을 적용한 수익률을 나타내며, 또한 그 다음날인 9월 6일에 KOSPI200 종목들 중에서 매수와 매도를 추천하는 종목을 나열한 예이다.

4. 결론 및 향후계획

본 연구는 유전 알고리즘과 기술적 분석을 사용하여 국내 주식 시장에서 매수와 매도를 나타내는 종목을 추출하는 방법에 대해 다루었다. 신뢰도가 높은 기법인 Slow Stochastic와 MACD 지표를 각 종목에 가장 알맞도록 유전 알고리즘을 이용하여 각 지표에 가장 적합한 기간을 계산한 다음 기술적 분석을 통해 매수 매도 신호가 나타난 종목들 중에서 가장 상위에 나타나는 종목을 추천한다.

한국 주식 시장에서 기술적 분석이 효과적으로 작용할 수 있는 안정적인 종목들을 고르기 위해 KOSPI 200의 종목으로 한정하여 실험하였으며, 이 종목들 중에서 유전 알고리즘을 적용하기 힘든 상장기간이 부족한 신규 종목들은 배제 하였다. 실험 결과 KOSPI 200 대부분의 종목들이 우수한 수익률을 보이고 있어 유전 알고리즘과 기술적 분석을 통한 투자기법이 긍정적인 가능성이 있음을 확인할 수 있었다.

본 연구는 종가에 대해서만 실험하여 실시간이 아닌 항상 하루 전의 주식 시장을 분석하여 추천하기 때문에 장기 투자에 적합하다. 본 연구 내용을 보다 실시간에 근접