

외환거래에서 의사결정나무와 그래디언트 부스팅을 이용한 수익 모형 연구[†]

정지현¹ · 민대기²

¹ 덕성여자대학교 통계학과 · ² 덕성여자대학교 정보통계학과

접수 2012년 12월 12일, 수정 2013년 1월 7일, 게재확정 2013년 1월 18일

요 약

외환차액거래는 국제외환 시장에서 외국의 통화를 거래하는 것으로 현물시장에서 이뤄지는 장의 통화선물 거래를 의미한다. 외환차액거래 데이터를 이용하여 의사결정나무와 그래디언트 부스팅 방법을 이용한 수익모형을 비교하였다. 금융시장의 예측을 위해 사용되고 있는 시계열분석과 같은 방법들은 장기간의 예측 모형을 설명하기에 장점이 있지만, 과동이 많고 짧은 시간에 가격이 급변하는 외환시장을 예측하기에는 한계가 있다. 따라서 본 논문에서는 단기간 즉 1, 3, 5분에서 외환시장의 수익구조를 의사결정나무와 앙상블기법의 하나인 그래디언트 부스팅으로 비교하여 매수, 매도거래 시 수익을 만들기 위한 규칙을 연구하였다.

주요용어: 그래디언트 부스팅, 앙상블, 외환차액거래, 의사결정나무.

1. 연구배경 및 목적

금융시장에서 많이 활용되고 있는 시계열분석과 같은 방법들은 장기간의 예측 모형을 설명하기에 장점이 있지만, 과동이 상대적으로 많은 단기간 급변하는 시장을 예측하기에는 한계가 있다. 따라서 데이터마이닝 (data mining) 분야에서 가장 많이 활용되는 의사결정나무 (decision tree)와 그래디언트 부스팅기법 (gradient boosting)을 적용하여 단기간 외환차액 거래를 예측하고 분석하였다. 그래디언트 부스팅은 앙상블기법의 하나로 예측력이 매우 우수한 기법으로 본 연구에서는 외환금융시장의 수익구조를 의사결정나무와 연구하였다. 앙상블기법은 여러 모형을 같이 사용하여 예측력을 높이는 방법으로, 이에 관한 다양한 연구는 최근에 Kim (2012) 등에 의해서도 볼 수 있다. 분석 자료는 외환종목 중 가장 거래량이 많고 활발한 EUR (Euro) / USD (US Dollar) 종목의 1분, 3분, 5분의 실시간 자료를 수집하여 의사결정나무와 그래디언트 부스팅으로 각각 30번 분석한 결과를 근거로 분리규칙을 만들어 비교하였다. 각 방법으로 만들어진 분리규칙은 이에 대한 일관성, 예측에 대한 오 분류율, 그리고 새로운 데이터에 적용하여 얻어진 수익을 통하여 방법에 대한 안정성과 적합도를 평가하였다.

[†] 위 연구는 2011 덕성여자대학교 교내연구비를 지원받았음.

¹ (132-714) 서울특별시 도봉구 근화교길19, 덕성여자대학교 통계학과 대학원, 석사.

² 교신저자: (132-714) 서울특별시 도봉구 근화교길19, 덕성여자대학교 정보통계학과, 교수.

E-mail: dkmin@duksung.ac.kr

2. 연구내용

본 연구에서 사용한 외환차액 거래 데이터는 FOREX (foreign exchange)라 불리는 국제외환 시장에서 개인이 직접 외국의 통화를 거래하는 현물 시장으로 장외 국외통화 선물거래를 의미한다. 외환차액 거래는 현재 세계에서 가장 자금의 흐름이 많은 금융시장으로 전 세계 주식시장 하루 거래량의 100배 이상의 규모이다. 또한, 시간에 구애받지 않고 24시간 거래할 수 있고, 유동성이 풍부하고, 안정성이 높다. 외환 차액 거래데이터는 장기적인 면에서 추세를 가지고 반복이 되므로 과거의 추세 분석과 보조지표 분석을 통하여 예측할 수 있으나, 단기간에는 거래 당사자들의 거래심리와 매번 발표되는 경제지표의 영향으로 급변하는 경우가 많아 실제로는 예측이 매우 힘들다. 본 연구에서는 변동하는 가격을 예측하는 접근법보다 비교적 파동이 적은 보조지표를 활용하여 매수, 매도 경우의 진입규칙에 관한 연구로 접근하였다. 외환거래에서 수익을 만들기 위한 필요조건은 상승이나, 하락의 추세가 지속하여야 한다. 가격상승이 예측되면 싼값에 기준통화를 매수하여 비싼 값에 팔아 수익을 만들고, 가격하락이 예측되면 기준통화를 매도하고 상대통화를 매수하여 수익을 내는 것이다. 그러나 가격이 항상 파동을 만들면서 즉 상승과 하락을 반복해가면서 추세를 지속하기 때문에 언제까지 추세가 지속할 것인가를 예측하기가 어렵다. 그러므로 본 연구에서는 가격, 추세, 강도를 나타내는 기술적 지표를 설명변수로 의사결정나무와 그레디언트 부스팅 방법을 이용하여 매수, 매도에 적절한 시기를 찾아내는 규칙을 발견하고자 한다.

2.1. 자료설명

분석 자료는 EUR/USD 종목의 1분, 3분, 5분 데이터를 사용하였고, EUR가 기준 통화이고, USD가 상대통화이다. 따라서 매수는 EUR를 매수하고 USD를 매도한다는 의미이고 매도는 EUR를 매도하고 USD를 매수한다는 의미이다. 이 자료는 미국의 FXCM이라는 미국 최대의 선물브러커회사의 자료를 국내 이트레이드회사의 온라인 거래를 통해서 얻은 실시간 자료이다.

2.1.1. 변수설명

외환거래의 모형을 위해 아래 Table 2.1의 보조지표를 설명변수로 분석에 사용하였다. 선택된 보조지표는 외환거래에서 차트에 대한 기술 분석으로 가장 많이 사용되는 것으로 MACD (moving average convergence & divergence), SS (stochastic slow), RSI (relative strength index) 와 가격변동 폭을 나타내는 BW (Bollinger bandwidth) 와 강도의 세기를 나타내는 ADX (average directional index), CCI (commodity channel index) 등이 있다.

Table 2.1 Variables description

variable	description	
Stochastic	momentum oscillator that measures overbought and oversold conditions over 80: overbought, below 20: oversold	$\%K = \frac{\text{Current Close} - \text{Lowest Low}}{\text{Highest High} - \text{Lowest Low}} * 100$
MACD (Moving Average Convergence & Divergence)	designed to identify trend changes BUY: MACD rises above its signal line SELL: MACD falls below its signal line	MACD line: 12-day EMA - 26-day EMA Signal line: 9-day EMA of MACD Line histogram: MACD-signal
RSI (Relative Strength Index)	momentum oscillator that measures the speed and change of price movements over 70: overbought, below 30: oversold	$RSI = 100 - \frac{100}{1 + RS}$ RS: Average Gain / Average Loss
DI+, DI- (Directional Movement Index)	shows the market tendencies BUY: when DI+ line crosses above the DI- SELL: when DI+ line crosses below the DI-	The DMI is the ratio of exponential moving averages of the greater of the upward (U) and downward (D) price movements, and the true range (TR). Calculation: $U = H_n - H_{n-1}$ $D = L_n - L_n$ $TR = (H_n - L_n) - (H_{n-1} - L_{n-1}) - (L_{n-1} - L_n)$ $EMAUP = EMAU_{n-1} + \frac{2}{(n+1)} * (U_n - EMAU_{n-1})$ $EMADOWN = EMAD_{n-1} + \frac{2}{(n+1)} * (D_n - EMAD_{n-1})$ $EMATR = EMATR_{n-1} + \frac{2}{(n+1)} * (TR_n - EMATR_{n-1})$ $+DI = EMAUP / EMATR$ $-DI = EMADOWN / EMATR$
ADX (Average Directional Movement Index)	momentum indicator that measures the strength of a prevailing trend if $ADX \geq 30$, the trend is confirmed, either up or down	$ADX = \frac{\sum[(+DI - (-DI)) / (+DI + (-DI))], N]}{N}$ N: the number of periods used in the calculation
CCI (Commodity Channel Index)	measures the position of price in relation to its moving average over 100: overbought, below 100: oversold	$CCI = \frac{\text{Price} - MA}{(0.015 * D)}$ CCI = (MA: moving average D: normal deviation
Bandwidth	measures the percentage difference between the upper band and the lower band	$(\text{Upper Band} - \text{Lower Band}) / \text{Middle Band}$



Figure 2.1 One minute candle chart of foreign exchange (EUR/USD)

2.1.2. 목표변수 설정

목표변수를 어떻게 설정할 것인가를 결정하는 것이 가장 어려운 문제 중의 하나였다. 왜냐하면, 단기 차트에서는 가격변동을 캔들 차트의 작은 시간단위로 나타내기 때문에 변동 폭이 아주 심하여 어떤 경우에 매수, 매도에 진입할 것인가를 정하는 것은 매우 힘든 일이다. 그래서 일반 거래인들 사이에는 가격의 움직임을 보고 거래를 시도하면 가격이 꼭 그 정반대로 움직인다고 말할 정도이다. 매수나 매도에 적당한 경우는 아주 선명하게 추세를 예측할 수 있는 경우이다. 이런 경우를 설정하기 위하여 단조 증가나 단조감소를 하는 보조지표를 활용하였다. Figure 2.1에서 두 번째 그래프는 추세의 강도를 나타내는 ADX와 ADX의 5일 이동평균선 나타내는 것이다. 가격의 흐름을 나타내는 위의 그래프와 동시에 비교하면 ADX가 이동평균선을 위로 통과할 때와 아래로 통과 할때가 각각 상승 강세, 하락 강세로 설정되어 핑크색과 푸른색으로 구간이 표시됨을 알 수 있다. 이렇게 강세구간 즉 수익을 만들 수 있는 거래 가능 구간을 목표변수로 설정하였다. 설정된 구간은 DI (directional index) 지표를 이용하여 상승이 하강보다 클 경우, 즉 DI+ (directional index plus)의 값이 DI- (directional index minus)의 값보다 큰 구간을 상승구간으로 지정하고, 반대인 경우를 하락구간으로 지정하여 구간별 수익을 계산하였다. 이때 구간의 수익이 3pip 이상일 경우 목표변수를 1로 설정하고, 3pip 미만일 경우 목표변수는 0으로 정하였다. 수익계산은 매수, 매도에서 평균값에서 진입하고 종가에서 청산하는 방식으로 계산하였고, 거래비용이나 수수료 등은 고려하지 않았다. 각 통화 사이의 값의 변화를 측정하는 단위를 pip이라고 한다. EUR/USD에서 1.2250에서 1.2251 상승하면 1pip이 변동하였다고 말한다. 통화쌍마다 pip 단위의 가치기준은 다르나 본 연구에서는 EUR/USD를 사용했기 때문에 1 pip의 가치를 \$10와 동일하다고 보면 된다. 상승구간 시작 시점을 매수 진입, 하락구간의 시작 시점을 매도 진입으로, 그 시점에서 보조지표들의 값을 설명변수의 값으로 설정하였다.

2.2. 연구진행절차

5분 이하의 단기 외환 금융시장의 수익모형분석을 위해서 2011년 10월부터 2012년 10월까지 13개월 안에서 임의로 충분히 큰 크기의 1분, 3분, 5분 자료를 수집하였다. 수집된 자료에서 매수나 매도시기로 가능한 경우가 목표변수로 설정되었다. Table 2.2는 각 시간대에서 목표변수의 빈도수를 표시하였다.

Table 2.2 Frequencies of target variable

Data	Target variable	Frequencies of bullish periods	Frequencies of bearish periods
1Minute	0	4446	4443
	1	905	888
3Minute	0	2687	2650
	1	926	913
5Minute	0	1586	1575
	1	653	621

SAS E-Miner 7.1의 의사결정나무와 그래디언트 부스팅을 분석방법으로 이용하였다. 전반적인 연구 진행과정은 Figure 2.2와 같이 1분, 3분, 5분 데이터에서 각각 무작위로 관측치 700개를 추출한 뒤, 의사결정나무와 그래디언트 부스팅을 이용해 모형을 설정하는 과정을 30번 반복하였다. 분리규칙을 생성한 후 생성된 분리규칙을 새로운 시간대의 데이터에 적용하여 수익성을 검증하였다. 그래디언트 부스팅은 예측결과에 다시 의사결정나무 방법을 적용해 분리규칙을 생성하였다.

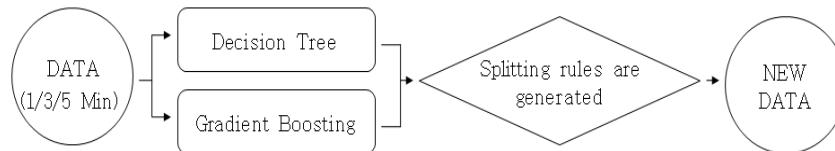


Figure 2.2 Process of study development

2.3. 연구방법

2.3.1. 의사결정나무

외환 금융시장의 모형을 위해 사용한 의사결정나무는 의사결정 규칙을 나무구조로 도표화 하여 관심 대상이 되는 집단을 분류, 예측하는 데이터마이닝의 대표적인 분석법이다. Kohavi 등 (1996)에 의하면 의사결정나무는 분석과정의 이해와 설명이 쉽다. 또한, 목표변수의 설명에 중요한 설명변수의 파악이 용이하고, 모형에 대한 가정이 필요 없다. 하지만 분리기준에 따라 다른 결과를 가져올 수 있으므로 상위단계의 분리에서 오류가 발생하였을 때, 하위단계의 분리에도 영향을 미치게 되어 신뢰할 수 없는 결과를 생성할 가능성이 있다. 또, 분석용 데이터에 크게 의존하여 새로운 자료의 예측이 불안정하다. Drucker과 Cortes (1996)에 의하면 이런 단점을 보완하고자 그래디언트 부스팅 알고리즘을 적용하면 정확한 예측을 할 수 있는 분류자를 형성하는데 매우 효과적이다.

2.3.2. 그래디언트 부스팅

그래디언트 부스팅은 Friedman (2001)에 의해 제안된 방법으로 단일 분류자를 사용하는 의사결정나무에 비해 여러 분류자들의 예측을 종합함으로써 분류의 정확성을 향상하는 앙상블기법의 하나이며 본 연구에서는 분류 정확도가 높은 최적의 모형을 찾기 위해 그래디언트 부스팅을 적용한다. Kim 등 (2005)에 의하면 부스팅은 SVM (support vector machine)에 비하여 예측력은 비슷하나 안정성이 결여된다고 하였다. 또한, Kim 등 (2012)에 의하면 트리의 적정크기를 통하여 모형의 예측력을 높일 수도 있다고 하였으나 본 연구에서는 고려하지 않았다. 부스팅은 기본 분류자가 분류하기 어려운 자료에 집중하도록 자료들의 분포를 변경하는데 사용되는 반복적인 절차이다. 그 중 그래디언트 부스팅은 분석용 데이터 관측값의 가중치가 동일한 상태에서 시작되어 형성된 분류자에 의해 오 분류된 관측값은 다음 관측값에 높은 가중치를 주게 되지만, 정 분류된 관측값은 낮은 가중치를 주는 과정을 반복함으로써 최종 분류자를 형성 한다. 그 과정은 아래 Figure 2.3과 같다. Figure 2.3에서 D 는 예측을 하고자 하는 학습 데이터의 원본이며, D_1, D_2, \dots, D_t 는 학습 데이터 원본 D 에서 생성한 데이터 집합이다. C_1, C_2, \dots, C_t 는 예측을 하기 위한 분류자이며 C^* 는 그래디언트 부스팅을 통하여 최종적으로 생성된 분류자이다.

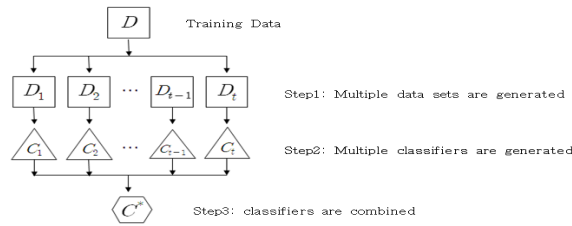


Figure 2.3 Logical developments of gradient boosting

2.4. 모형비교

의사결정나무와 그래디언트 부스팅으로 생성한 모형의 비교를 위해서 분리된 규칙의 일관성과 모형의 오 분류율을 비교하였다. 또, 의사결정나무와 그래디언트 부스팅으로 생성한 분리규칙을 새로운 데이터에 적용하여 수익을 내는 횟수, 평균 수익, 수익 구간이 연속되는 시간의 평균을 비교하여 단기 외환 금융시장에서 의사결정나무와 그래디언트 부스팅으로 생성한 모형을 평가하였다.

3. 연구결과

3.1. 분리규칙생성

Table 3.1은 외환차액거래 1분 데이터에서 30번의 반복실험을 통하여 의사결정나무와 그래디언트 부스팅 모형을 각각 30개 생성한 후 분리기준에 따라 절단 점과 오 분류율의 평균을 계산한 결과이다. 매수에서 그래디언트 부스팅을 이용해 분리하였을 때 첫 번째 분리가 CCI, 두 번째 분리가 BW에 의해 되는 경우가 30번 중 8번으로 가장 많이 발생하였다. 의사결정 나무를 이용하였을 때 첫 번째 분리는 BW, 두 번째 분리는 CCI의 경우가 30번 중 4번 발생하였다. 매도는 그래디언트 부스팅을 이용해서 분리하였을 때, 첫 번째 분리가 BW, 두 번째 분리가 CCI에 의해, 되는 경우의 빈도가 30번 중 9번 발생하였다. 의사결정나무를 이용해서 분리하였을 때, 첫 번째 분리는 BW, 두 번째 분리가 CCI에 의해 발생빈도가 30번 중 7번 발생하였다. 1분 차트의 결과를 종합해보면 그래디언트 부스팅의 경우에는 순서에 상관없이 BW, CCI 조합의 규칙이 매수, 매도를 합친 60번 중에서 28번이 발생했고, 의사결정나무는 19번 발생하였다. 오 분류율의 평균을 비교해 보면 전체 평균이 그래디언트 부스팅인 경우 0.27, 의사결정의 경우는 0.28로 차이가 없어 보인다.

Table 3.1 Splitting rules of one minute chart

Method	Data	Splitting rule		Frequencies	Misclassification rate			
		First	Second					
Gradient boosting	Buy	CCI	≥ 140.8518	BW	≥ 0.000413	8/30	0.251316	
		BW	≥ 0.0006222	CCI	≥ 125.8523	7/30	0.217575	
	Sell	BW	RSI	≥ 56.0825	CCI	< -110.325	4/30	0.307172
			CCI	≥ 0.000617	RSI	< 41.663	9/30	0.26504
		CCI	RSI	< -138.153	BW	≥ 0.000475	5/30	0.275655
			BW	≥ 0.000475	CCI	≥ 153.4996	4/30	0.309828
Decision tree	Buy	BW	≥ 0.0007	RSI	≥ 66.62875	4/30	0.305912	
		CCI	≥ 153.2393	BW	≥ 0.000267	4/30	0.281964	
	Sell	BW	≥ 0.000594	CCI	< -133.5040	3/30	0.259454	
		CCI	< -132.203	RSI	< 31.8583	7/30	0.245958	
		RSI	< 31.8583	BW	≥ 0.00056	3/30	0.336553	
		CCI	< -132.203	BW	≥ 0.00056	5/30	0.255854	

Table 3.2는 외환차액거래 3분 데이터에서 30번 반복실험을 의사결정나무와 그래디언트 부스팅 모형으로 시행한 결과이다. 1분에서와 마찬가지로 그래디언트 부스팅을 이용한 경우에 BW와 CCI의 조합이 매수, 매도를 포함하여 전체 60번 중 26번 결과가 나왔다. 반면 의사결정나무를 이용한 경우에는 전체 60번 중 14번에서 BW와 CCI 결과가 나왔다. 오 분류율 결과의 평균값을 비교하면 앞의 1분 자료에서는 차이가 없어 보였으나 3분 자료에서는 그래디언트 부스팅이 0.25, 의사결정나무가 0.31로 확인한 차이가 있어 방법에 대한 우수성을 알 수 있었다.

Table 3.2 Splitting rule of three minute chart

Method	Data	Splitting rule			Frequencies	Misclassification rate	
		First	Second				
Gradient boosting	Buy	BW	≥ 0.001106	CCI	≥ 138.8161	8/30	0.22803
				RSI	≥ 59.7375	4/30	0.319847
				BW	≥ 0.000925	4/30	0.295255
	Sell	CCI	≥ 134.452	BW	≥ 0.000975	4/30	0.242185
		BW	≥ 0.001029	CCI	< -149	8/30	0.254056
		CCI	< -144.715	BW	≥ 0.000733	6/30	0.230393
		Stochastic	≥ 10.885	3/30	0.197683		
Decision tree	Buy	BW	≥ 0.001071	CCI	≥ 179.1446	6/30	0.343245
				DI+	≥ 32.155	2/30	0.282222
				BW	≥ 0.0008	3/30	0.299699
	Sell	CCI	≥ 149.4094	MACDOSC	≥ 0.0001	3/30	0.331215
		RSI	≥ 56.92743	BW	≥ 0.00106	5/30	0.275922
		CCI	< -132.944	CCI	< -171.845	3/30	0.316803
		Stochastic	≥ 10.07167	3/30	0.310578		

Table 3.3은 외환차액거래 5분 데이터에서 결과이다. 앞에서 살펴본 1분, 3분 결과와 같이 BW, CCI 조합의 규칙이 매수, 매도를 포함한 60번 중 22번에서 나왔다. 반면 의사결정나무를 이용한 분석은 60번 중 10번만 나왔고 그 밖에 다른 규칙이 발견되어 자료에 따라 분리규칙이 민감하게 좌우된다는 것을 알 수 있었다. 5분 자료에서 정확도면에서 비교하기 위한 오 분류율의 평균은 그래디언트 부스팅은 0.26인 반면 의사결정나무는 0.36으로 나타났다.

Table 3.3 Splitting rules of five minute chart

Method	Data	Splitting rule			Frequencies	Misclassification rate	
		First	Second				
Gradient boosting	Buy	BW	≥ 0.001331	CCI	≥ 109.0177	10/30	0.295681
				BW	≥ 0.00152	5/30	0.229449
	Sell			CCI	≥ 122.9826	3/30	0.247951
		CCI	< -150.379	BW	≥ 0.00105	12/30	0.25624
Decision tree	Buy	RSI	≥ 61.31542	CCI	≥ 142.3309	4/30	0.363265
				CCI	≥ 131.6046	3/30	0.35102
				BW	≥ 0.001633	3/30	0.434695
	Sell	BW	≥ 0.001671	CCI	≥ 131.6046	3/30	0.35102
		CCI	≥ 178.146	BW	≥ 0.001633	3/30	0.434695
		CCI	< -154.307	BW	≥ 0.001025	4/30	0.285714
		DI-	< 40.92167	3/30	0.361905		
		RSI	< 38.35563	MACDOSC	< -0.00006	3/30	0.338333

Table 3.4는 1분, 3분, 5분 데이터에서 의사결정나무와 그래디언트 부스팅을 이용하여 분석한 결과 30개 중 가장 많이 발견된 분리규칙과 결과이다. 그래디언트 부스팅의 경우 매수와 매도 두 경우 모두 1분, 3분, 5분 자료에서 CCI와 BW에 의해 분리가 되는 경우가 가장 많이 발생하여 분리규칙의 안정성을 알 수 있지만, 의사결정나무의 매수는 1분, 3분에서는 BW와 CCI에 의해 분리가 되었지만 5분에서는 RSI와 CCI에 의해 분리가 되어 1분, 3분, 5분의 분리규칙이 일정하지 않다는 것을 알 수 있었다. 이를 통하여 의사결정나무가 그래디언트 부스팅 보다 절단 값에 많이 좌우되어 분리규칙이 일정하지 않음

을 알 수 있었고 변동이 심한 자료를 분석할 때 적용하기가 적합하지 않다는 것을 알 수 있었다. 정확도에서도 Table 3.4 결과를 근거로 두 방법에 대한 오 분류비율의 평균은 그래디언트 부스팅이 0.26 표준편차는 0.02인 반면 의사결정나무의 평균은 0.30, 표준편차는 0.04로 확실한 차이가 있어, 방법에 대한 우수성을 확인 할 수 있었다.

Table 3.4 Summary of splitting rules

Method	Data	Splitting rule		Frequencies	Misclassification rate	
		First	Second			
Gradient boosting	Buy	1Min	CCI ≥ 140.8518	BW ≥ 0.000413	8/30	0.251316
		3Min	BW ≥ 0.001106	CCI ≥ 138.8161	8/30	0.22803
		5Min	BW ≥ 0.001331	CCI ≥ 109.0177	10/30	0.295681
	Sell	1Min	BW ≥ 0.000617	CCI < -110.325	9/30	0.26504
		3Min	BW ≥ 0.001029	CCI < -149	8/30	0.254056
		5Min	CCI < -150.379	BW ≥ 0.00105	12/30	0.25624
Decision tree	Buy	1Min	BW ≥ 0.0007	CCI ≥ 153.4996	4/30	0.305912
		3Min	BW ≥ 0.001071	CCI ≥ 179.1446	6/30	0.343245
		5Min	RSI ≥ 61.31542	CCI ≥ 142.3309	4/30	0.363265
	Sell	1Min	BW ≥ 0.000594	CCI < -133.5040	7/30	0.245958
		3Min	CCI < -132.944	BW ≥ 0.00106	5/30	0.275922
		5Min	CCI < -154.307	BW ≥ 0.001025	4/30	0.285714

3.2. 모형평가

두 연구방법의 예측력을 평가하기 위하여 오 분류율을 비교하였다. Table 3.5는 의사결정나무와 그래디언트 부스팅을 이용하여 30회 분석을 시행한 결과다. 이미 발표된 연구결과에서 알고 있듯이 모형들의 오 분류율의 평균 표준편차를 비교해본결과 그래디언트 부스팅을 이용하여 생성된 모형들에 비해 의사결정나무에 의해 생성된 모형들의 오 분류율 평균이 훨씬 높았다. 이를 통해서 그래디언트 부스팅방법의 의사결정나무에 대한 예측력의 우수성을 확인할 수 있었다.

Table 3.5 Comparison of misclassification rate

Data	Buy				Sell			
	Gradient boosting		Decision tree		Gradient boosting		Decision tree	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
1Min	0.281055	0.05937	0.300215	0.06121	0.28187	0.053734	0.29235	0.099208
3Min	0.270015	0.069443	0.33008	0.050087	0.254951	0.069005	0.318335	0.048442
5Min	0.265576	0.07211	0.344991	0.068503	0.276398	0.051066	0.336751	0.042828

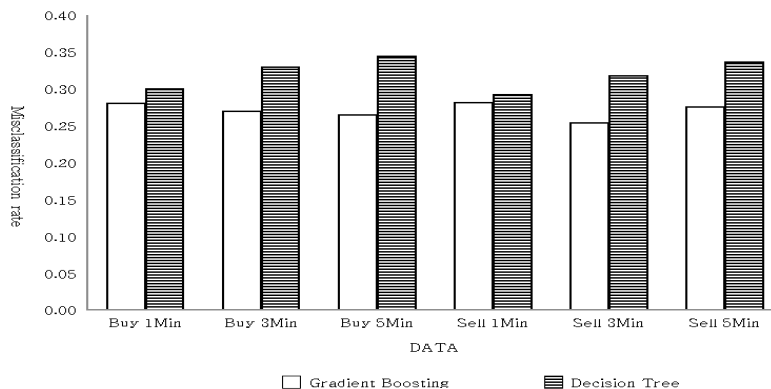


Figure 3.1 Comparison of misclassification rate

3.3. 수익비교를 통한 모형검증

의사결정나무와 그래디언트 부스팅을 이용하여 생성한 모형을 검증하기 위하여 수익비교를 하였다. 분석방법을 통하여 발견한 분리규칙을 아래 Table 3.6에 언급한 시간 데이터에 적용하여 수익성 평가를 하였다.

Table 3.6 Test data description

Data	Periods	Observations
1Min	2012/10/31 11:20:00 - 2012/11/01 13:46:00	1586
3Min	2012/10/24 07:42:00 - 2012/11/01 13:45:00	3000
5Min	2012/10/23 15:10:00 - 2012/11/01 13:45:00	2000

Table 3.7은 의사결정나무와 그래디언트 부스팅의 분리규칙을 새로운 데이터에 적용하여 계산한 3pip 이상 빈도/총빈도, 평균 pip, 연속 상승, 하락 구간의 평균이다. Table 3.7과 같이 1분 매수, 매도, 3분 매수, 5분 매수, 매도에서 그래디언트부스팅으로 생성한 분리규칙이 의사결정나무로 생성한 분리규칙보다 3pip 이상 수익을 내는 횟수와 평균 수익, 수익 구간이 지속하는 평균시간에서 우월하다.

Table 3.7 Evaluation of profits depend on splitting rules

		Method	Over 3pips Percents	Average Pips	Average Pips During the Period
1Min	Buy	Gradient Boosting	4/25=16%	0.96	3.08
		Decision Tree	1/17=5.88%	0.5882353	2.0588235
	Sell	Gradient Boosting	4/25=16%	0.99333	3.2
		Decision Tree	2/23=8.7%	0.5420290	2.5217391
3Min	Buy	Gradient Boosting	8/39=20.51%	1.5384615	2.5641026
		Decision Tree	4/26=15.38%	1.0769231	1.6923077
	Sell	Gradient Boosting	8/38=21.05%	1.9614035	2.4210526
		Decision Tree	9/45=20%	2.144444	2.666667
5Min	Buy	Gradient Boosting	8/29=27.59%	3	3.4482759
		Decision Tree	8/40=25.58%	1.9534884	2.5116279
	Sell	Gradient Boosting	11/31=35.48%	3.3752688	2.5483871
		Decision Tree	11/31=35.48%	3.3720430	2.4193548

4. 결론

본 연구는 5분 이하에서 거래되는 단기 외환시장에서 수익을 내는 거래 진입 규칙을 찾기 위하여 의사결정나무와 그래디언트 부스팅을 활용하였다. 의사결정나무는 절단 점에 따라 분리가 결정된다는 약점이 있고, 단기 외환시장은 변동이 많아서 안정적 규칙을 찾기 위하여 그래디언트 부스팅 방법을 적용하여 비교하였다. 30개의 자료를 통해서 얻어진 분리규칙을 비교하여 그래디언트 부스팅 결과의 안정성을 알 수 있었다. 또한, 오 분류율을 비교하였을 때, 그래디언트 부스팅 보다 의사결정나무를 이용하여 생성한 모형의 정확도가 낮음을 확인할 수 있었다. 두 분석방법을 활용한 모형의 분리규칙을 새로운 데이터에 적용하였을 때, 그래디언트 부스팅으로 생성한 모형의 규칙이 의사결정나무로 생성한 규칙보다 수익성이 우수하였다. 따라서 변동이 심하고 과동이 많이 존재하는 단기 외환시장에서 그래디언트 부스팅 분석방법의 우월성을 알 수 있었다.

참고문헌

- Bauer, E. and Kohavi, R. (1999). An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine Learning*, **55**, 1-38.
- Drucker, H. and Cortes, C. (1996). Boosting decision trees. *Neural Information Processing*, **8**, 470-485.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, **29**, 1189-1232.
- Kim, A., Kim, J. H. and Kim, H. (2012). The guideline for choosing the right-size tree for boosting algorithm. *Journal of the Korean Data & Information Science Society*, **23**, 949-959.
- Kim, Y. D., Kim, K. H. and Song, S. H. (2005). Comparison of boosting and SVM. *Journal of the Korean Data & Information Science Society*, **16**, 999-1012.
- Kohavi, R., Sommerfield, D. and Dougherty, J. (1996). Data mining using MLC. *International Journal on Artificial Intelligence Tools*, **6**, 234-245.
- Kim, S. H., Cho, D. H. and Seok, K. H. (2012). Study on the ensemble methods with kernel ridge regression. *Journal of the Korean Data & Information Science Society*, **23**, 375-383.

The study of foreign exchange trading revenue model using decision tree and gradient boosting[†]

Ji Hyeon Jung¹ · Dae Kee Min²

¹Department of Statistics, Duksung Women's University

²Department of Information & Statistics, Duksung Women's University

Received 12 December 2012, revised 7 January 2013, accepted 18 January 2013

Abstract

The FX (Foreign Exchange) is a form of exchange for the global decentralized trading of international currencies. The simple sense of Forex is simultaneous purchase and sale of the currency or the exchange of one country's currency for other countries'. We can find the consistent rules of trading by comparing the gradient boosting method and the decision trees methods. Methods such as time series analysis used for the prediction of financial markets have advantage of the long-term forecasting model. On the other hand, it is difficult to reflect the rapidly changing price fluctuations in the short term. Therefore, in this study, gradient boosting method and decision tree method are applied to analyze the short-term data in order to make the rules for the revenue structure of the FX market and evaluated the stability and the prediction of the model.

Keywords: Decision tree, ensemble, FX margin, gradient boosting.

[†] This paper was supported by Duksung Women's University in 2011.

¹ Master, Department of Statistics, Duksung Women's University, Seoul 132-714, Korea.

² Corresponding author: Professor, Department of Information & Statistics, Duksung Women's University, Seoul 132-714, Korea. E-mail: dkmin@duksung.ac.kr