

투자전략 보고서의 제목이 주가 예측에 미치는 영향: 텍스트마이닝 중심으로

How the Title of Investment Strategy Report Affects Stock Price Forecast:
Using Text Mining Method

장준규 · 이규현 · 이준기[†]

연세대학교 정보대학원

요 약

재무분석가의 투자전략 보고서는 전문가의 금융정보 분석으로 사람들간의 금융 정보 격차를 줄일 수 있게 만들어 줄 수 있다. 하지만 여러 다른 인센티브로 말미암아 투자전략 보고서는 왜곡의 소지가 있을 수 있다. 만약, 투자전략 보고서 제목만으로 예측에 미치는 영향을 파악할 수 있다면, 예측력이 높은 투자전략 보고서를 분별할 수 있게 된다. 이에 본 연구는 재무분석가의 투자전략 보고서의 제목과 예측의 관계를 파악하려 한다. 텍스트 마이닝을 사용하여, 국내 재무분석가의 투자전략 보고서 제목으로부터 연구의 중요변수인 투자의견을 추출하고 모형을 구축하여 재무분석가의 예측 정확도와 예측 달성도를 측정하였다. 분석 결과, 강한 매수의견과 매도의견이 제목에 있을수록 재무분석가의 예측 정확도와 예측 달성도가 높아짐을 검증하였다. 본 연구 결과가 투자자에게 더 정확한 투자전략 보고서를 판단하는 기준을 제시하기를 바라고 빅데이터를 통한 분석 연구에 시사점을 주길 기대한다.

- 중심어 : 재무분석가 예측, 텍스트마이닝, 한글 형태소 분석, 투자전략 보고서, 투자의견

Abstract

There are various investment strategy reports available online, prepared by many financial analysts. If the correlation between the title of the report and analyst forecast can be found, we can tell from the title whether analyst' forecast will be reliable or not. The objective of this study is to see the correlation between the title of analyst investment strategy report and the actual result of forecast by using the Text Mining technique. The result of actual analysis showed that "strong buy and sell call" appeared in the title lead the higher accuracy of analyst forecast and fulfillment ratio. The results that potential investors can get better information by reading the title of the analyst report. We hope that this study could be the basis for new methodologies in this area.

- Keyword : Financial Analysts Forecast, Text Mining, Hangle Morpheme Analyzer, Investment Strategy Report, Investment Opinions

I. 서론

최근 정보기술의 발달로 사람들은 수많은 정보를 생성하고 공유하게 되었다. 이에, 과거에 정보 격차가 심했던 금융 투자정보의 경우에는 재무분석가가 온라인 상으로 투자전략 보고서를 제공함으로써, 정보의 비대칭성을 상당부분 해소할 수 있게 되었다. 하지만, 다양한 형태의 많은 투자전략 보고서가 온라인상에 존재하게 되었고, 정보를 습득하는 입장에서는 정보선택에 어려움이 따르고 있다. 이에 본 연구는 투자자에게 신호가 될 정보를 재무분석가의 전략보고서에서 찾아내고자 한다.

투자전략 보고서에 공시된 투자의견은 관계사, 고객사, 주간사 등의 외적 영향을 많이 받기에, 본 연구에서는 투자전략 보고서의 제목이 투자자에게 신호(Signal)가 될 것이라 추정하였으며, 분석의 용이함을 위해, 투자의견에 대해 강한 매수, 매수 그리고 매도의 투자의견으로, 총 세 가지로 나누어서 보려고 한다.

본 연구에서는 선행연구를 바탕으로 금융분야에 텍스트 마이닝을 적용하고자 하며, 재무분석가의 의도를 파악하기 위해, 재무분석가의 투자전략 보고서에 제시된 투자의견이 아니라, 외적 영향이 가장 최소화된 재무분석가의 투자전략 보고서의 제목에서 투자의견을 추출 할 것이다. 그렇게 추출된 투자의견이 재무분석가의 예측에 영향을 끼치는지를 보고자 한다. 재무분석가의 예측력의 측정방법은 Bradshaw and Brown[15]의 연구에서 사용한 2가지 측정방법을 채택하였다.

본 논문의 구성은 다음과 같다. 제II장에서는 선행 연구를 정리하며, 그것으로부터 도출된 가설을 설정할 것이다. 제III장에서는 본 연구에서 제안하는 재무분석가 투자전략 보고서 제목의 가치를 측정하는 모형을 소개하였다. 다음으로, 제IV장에서는 실증 분석 결과를 정리하였으며, 제V장에서는 투자자들이 활용할 수 있도록 연

구의 시사점 및 한계점에 대해서 논의하였다.

II. 개념적 배경

2.1 재무분석가와 투자의견에 관한 선행연구

재무분석가의 선행연구에서 살펴보면, 재무분석가들은 이익예측 정확도에 대해서 차별화된 능력이 있으며[15, 25, 28], 특히 과소·과대평가된 주식을 판별하는 능력이 있다[28].

재무분석가의 투자의견에 따라 획득할 수 있는 수익률 측정에서 연 4%의 초과수익이 나타났다[13]. 이런 차별화된 능력은 재무분석가와 투자자들은 서로 다른 정보를 보유하고 있기 때문이다. 이는 즉, 정보불균형이 존재한다[30].

An et al.[11]의 연구에서는 스타트업의 성공을 위한 투자유치가 중요하기 때문에 이를 위해 스타트업이 잠재적인 투자자들에게 줄 수 있는 신호와 소음(signal and noise)이 있다는 것을 밝혀내었다.

재무분석가의 투자전략 보고서는 투자자들에게 참고가 된다. 재무분석가의 투자의견이 주가에 반영되고 있음이 이를 증명한다. 재무분석가의 투자의견이나 목표주가를 상향 조정할 경우 주가가 유의하게 상승하였고, 하향 조정의 경우에는 주가가 유의하게 하락하였기 때문이다[2].

이러한 상황에서 재무분석가의 예측치나 투자의견은 여러 가지 환경에 의해서 제시된다. 주간사 회사의 재무분석가가 비주간사 회사의 재무분석가보다 더 낙관적인 예측을 하며[8], 소속 증권회사의 고객에 대한 관계로 인하여 투자의견이나 이익예측을 낙관적으로 한다[1, 20, 21, 30]. 따라서 재무분석가의 매수의견이 매도 의견보다 빈번하거나 매도나 강한 매도 투자의견이 거의 발표되지 않는다[12, 30]. 따라서 투자전략 보고서에 공시된 투자의견은 외적 영향을 많이 받는다. 따라서 본 연구에서는 투자전략 보고서의 제목이 투자자에게 신호(Signal)가 될 것이라 추정하였다.

텍스트 마이닝의 선행연구에서 살펴보면, 텍스트는 현실에서 정보를 표현하거나 교환하는 방법으로 가장 널리 사용하는 수단이다[29]. 텍스트 데이터의 형태는 비정형 데이터라고 불리며 이를 분석하여 이전에는 찾을 수 없었던 의미 있고 유용한 패턴을 추출하는 과정을 텍스트 마이닝이라 한다[19, 24]. 텍스트 마이닝의 선행연구는 컴퓨터 공학, 문헌정보학, 정보통신학, 공학 등 다양한 분야에서 시도 되고 있다[3]. 따라서 선행연구를 통해서 텍스트 마이닝이 다양한 분야에 접목 가능하다는 것을 알았다.

결론적으로 본 연구에서는 선행연구를 바탕으로 금융분야에 텍스트 마이닝을 적용하고자 하며, 재무분석가와 관계회사, 고객사, 주간사회사의 영향을 최소화하며, 재무분석가의 투자 의견을 명확히 판단하기 위해 재무분석가의 투자 전략 보고서에 제시된 투자 의견이 아니라, 재무분석가의 투자 전략 보고서의 제목에서 투자 의견을 추출 할 것이다. 그렇게 추출된 투자 의견이 재무분석가의 예측에 영향을 끼치는지를 보고자 한다. 재무분석가의 예측력의 측정방법은 Bradshaw and Brown[15]의 연구에 사용한 2가지 방법을 사용한다. 따라서 본 연구에서는 다음과 같은 가설을 설정한다.

- 가설 1: 투자전략 보고서의 제목의 의미에 따라 재무분석가의 예측 정확도에 영향을 끼친다.
- 가설 2: 투자전략 보고서의 제목의 의미에 따라 재무분석가의 예측 달성도에 영향을 끼친다.

2.2 자연어 처리(Natural Language Processing)

텍스트는 현실에서 정보를 표현하거나 교환하는 방법으로 가장 널리 사용하는 수단이다[29]. 최근에는 정보통신기술의 발달로 인해 인터넷 문화가 널리 확산되었으며 웹과 모바일에서 다양한

소셜미디어가 존재한다. 이러한 소셜미디어로부터 방대한 양의 텍스트 데이터가 생산된다. 이런 텍스트 데이터의 형태는 비정형 데이터라고 불리며 이를 분석하여 이전에는 찾을 수 없었던 의미 있고 유용한 패턴을 추출하는 과정을 텍스트 마이닝이라 한다[19, 24]. 기존의 데이터 마이닝에서 확장된방법인 텍스트 마이닝은 군집화, 연관 분석, 분류, 자연어 처리, 토픽분석 그리고 텍스트 범주화 등 다양한 형태의 기술을 종합적으로 활용한다 [3, 7, 23, 27]. 특히 텍스트 마이닝의 자연어 처리 기술은 차별화된 텍스트 마이닝만의 핵심 기술이다. 사람들이 일상적으로 사용하는 언어를 자연어라 말한다. 사람들은 같은 의미이지만 서로 다른 표현방식을 사용한다. ‘먹다’라는 동사의 의미를 ‘먹었다’, ‘먹었죠’, ‘먹습니다’, ‘먹고 있다’ 등의 표현 방식을 사용한다. 이런 자연어가 컴퓨터가 인식할 수 있도록 인공어(Artificial Language)로 변환시키는 것이 자연어 처리이라 한다.

III. 연구방법

3.1 데이터 수집

본 연구의 주요변수를 측정하기 위하여 FnGuide의 재무분석가의 투자전략 보고서의 제목을 사용하였다. 분석 도구로는 R 3.2.2와 SPSS 21을 이용했다.

첫 단계로는 데이터를 수집하였다. FnGuide에서 2014년 04월 01일부터 2015년 03월 31일까지 수집한 31970개의 재무분석가 투자전략 보고서의 제목을 형태소로 바꾸는 작업을 하였다. 수집기간 설정이유는 2013년 12월 결산이 2014년 03월 31일에 공시되며 이를 재무분석가가 예측할 때 참조하기 때문이다. 즉, 2014년 3월 31일 이후의 재무분석가의 투자전략보고서는 2013년 12월 결산(2014년 03월 31일 공시)을 참조하여 작성한다. 본 연구에서 기간을 1년 단위로 측정하기 때문에 2014년 12월 결산(2015년 03월 31

일 공시)자료를 이용할 시, 재무분석가의 투자 전략 보고서의 수집기간은 2015년 04월 01일부터 이다. 둘째로, 기업 중 다음 조건을 만족시키는 기업을 표본으로 선정하였다(<표 1> 참조).

- (1) 금융업 제외
- (2) 12월 결산일인 기업
- (3) FnGuide에서 재무분석가 예측치와 제목을 수집 가능한 기업
- (4) NewsKisvalue의 데이터베이스에서 기업 추가자료가 없는 기업 제외
- (5) TS2000 데이터베이스에 대상기업의 재무 자료가 없는 기업 제외
- (6) 공시발표일(2015.03.31)전 3개월 치 데이터 제외

조건 (1)과 (2)는 표본의 동질성을 위한 조건이고, 조건 (3)은 종속변수인 재무분석가의 예측자료와 본 연구의 중요변수를 도출하는데 사용되는 투자 전략 보고서의 제목을 수집하기 위한 조건이다.

조건 (4)와 (5)는 분석의 용이함을 위한 조건이며, 조건 (6)은 재무분석가의 예측치가 공시발

표일에 가까워질수록 예측 토대가 되는 정보의 량이 많아 상대적으로 과거에 투자전략 보고서를 발표한 재무분석가들보다 예측이 더 정확할 것이라 판단되었다. 따라서 이 조건은 결과의 왜곡을 피하기 위함이다.

본 연구에서는 위의 기준을 만족하여 최종적으로 선택된 표본은 10,318개이며 <표 1>에 표본추출과정을 정리하였다. 마지막으로 극단치가 결과에 미치는 영향을 통제하기 위해서, 연속변수에 대해 양극단 1%를 벗어나는 관찰치는 극단치로 보아 절사하였다.

3.2 투자전략 보고서 제목 분석

본 연구는 텍스트 마이닝인 자연어 처리 및 분석 과정이 필수적이다. 따라서 첫 번째 분석 단계로 R의 ‘KoNLP’ 패키지를 이용하여 형태소 분석을 진행하였다. 형태소는 의미의 최소단위로 더 이상 분리가 불가능한 작은 요소이다. 수집한 제목 중 예를 들어, ‘기업분석 SK텔레콤 최고의 배당주이자 단통법 최대 수혜주’를 명사형으로 형태소 분석을 하면 ‘기업’, ‘분석’, ‘S’, ‘K’, ‘텔레콤’, ‘최고’, ‘배당주’, ‘단’, ‘통’, ‘법’, ‘최대’, ‘수혜주’로 분석이 된다. 형태소 분석을 할 경우 어떠한 사전과 분석기를 쓰느냐에 따라서 결과가 달라진다.

다음 분석 단계로는 ‘RWeka’ 패키지를 이용하여 N-gram함수를 적용하였다. N-gram함수 적용시, uni-gram, bi-gram, tri-gram, quad-gram을 사용하였다. N-gram함수의 N은 한 문장 안의 분석단어 수를 의미하며, 한 문장 안에 하나의 분석단어가 있을 경우 uni-gram, 두 가지의 분석단어가 있으면 bi-gram이며, tri-gram, quad-gram또한 이러한 형태로 이름 붙여진다.

예를 들어, uni-gram을 적용할 경우 ‘최대’, ‘수혜주’로 나오는 반면 bi-gram은 ‘최대 수혜주’로 결과가 도출된다. 이것의 장점은 문장내의 단어의 순서와 위치를 알 수 있다는 것이다. 본 연구에서

<표 1> 표본추출과정

구 분	표본수
전체 2014. 04. 01~2015. 03. 31	31,970
FnGuide에서 수집한 자료에 목표 주가가 없는 것	(4,894)
FnGuide에서 수집한 자료에 제목이 없는 것	(723)
FnGuide에서 수집한 자료에 외국인투자를 상대로 한 글	(5,185)
FnGuide에서 수집한 자료가 요약 글일 경우	(2,042)
FnGuide에서 수집한 기업코드가 한국거래소에 없는 경우	(862)
NewsKisvalue에 수집 기간 내의 기업 추가자료가 없는 경우	(3,636)
TS2000에 해당 기업 재무자료가 없는 경우	(1,094)
공시발표일(2015. 03. 31)전 3개월치 데이터	(3,216)
최종 표본	10,318

N-gram함수를 적용 후, removeSparseTerms 함수를 이용해 빈도수가 희박한 단어조합은 분석에 불필요하다고 판단하여 제외시켰다. 그 결과, 68개의 단어의 조합이 도출되었다(<표 2> 참조).

‘최대_수혜’와 ‘최대_수혜_주’가 포함된 제목을 확인한 결과 동일한 의미를 가지기 때문에 둘 중 한가지 단어조합을 제외하였다. 이와 같은 단어(‘양호_실적’과 ‘양호실_적’)를 제외하였다. 또한 ‘실적’, ‘시장기_대치’, ‘하반기_실적’, ‘진행_중’의 단어조합은 해당 제목의 투자의견을 추출할 수 없다고 판단하여 제외시켰다. 왜냐하면 ‘시장기_대치’가 들어간 제목이 ‘시장기_대치_부합’와 ‘시장기_대치_상회’ 둘 중 어떤 것을 뜻하는지를 비교적 명확하지 않기 때문이다. 최종적으로 추출된 단어 조합 62개를 재무분석가의 투자전략 보고서를 참고하는 전문가 6명의 인터뷰를 통해, 강한 매수, 매수, 매도의견으로 범주화 하였다.

인터뷰의 목적은 제목의 단어 조합에 따른 투자의견을 도출하는 것이다. 인터뷰 기간은 2014

년 12월 12일이며 인터뷰 대상자는 투자를 할 경우, 재무분석가의 투자전략보고서를 참고하는 투자자로 국한하였다. 인터뷰 과정은 도출된 전체 단어 조합을 설문지에 나열 하였다. 그 후 인터뷰 대상자에게 각각의 단어 조합을 살펴보고 해당 단어 조합이 강한 매수의 투자의견이라 판단되면, 1로 표시하게 하였고 매수, 매도의견이라 판단될 때는 각각 2와 3을 표시하게 하였다. 인터뷰 후 본 연구에서는 인터뷰 대상자들이 표시한 숫자의 빈도수가 가장 높은 것을 해당 투자의견으로 판단하였다(<표 3> 참조). 예를 들어 ‘견조_실적’의 단어 조합의 결과로, 5명의 인터뷰 대상자가 강한 매수 투자의견이라 하였기 때문에 ‘견조_실적’은 강한 매수 투자의견으로 포함하였다.

결론적으로 본 연구의 분석 데이터로 활용할 세가지 말뭉치로 나누었다. 그 결과는 아래 <표 4>, <표 5>, <표 6>와 같다.

<표 2> 제목으로부터 추출한 단어 조합

개선_기대	부진_분기	실적	예상치_부합
개선_예상	부진_실적	실적_개선	장기_성장
개선_전망	부진_지속	실적_개선_지속	주주_환원
개선_지속	부합_분기	실적_달성	지속_전망
개선_추세	부합_실적	실적_리뷰	진행_중
견조_실적	분기_실적	실적_모멘텀	최대_수혜
기다_리	분기_실적_부진	실적_발표	최대_수혜_주
기대_부합	불확실성_해소	실적_부진	컨센서스_부합
기대치_부합	성장_모멘텀	실적_예상	컨센서스_상회
기대치_상회	성장세_지속	실적_전망	컨센서스_하회
기대치_하회	성주_목	실적_턴어라운드	컨퍼런스콜요약
단기_실적	소폭_하회	실적_호조	탐방_노트
대감_유효	수익성_개선	실적_회복	필요시_점
대치_부합	수혜_주	양호_실적	하반기_실적
대치_상회	시장기_대치	양호실_적	하반기_실적_개선
매수_기회	시장기_대치_부합	예상_부합	회복기_대
무난_실적	시장기_대치_상회	예상_수준	흑자_전환

〈표 3〉 인터뷰 결과표

	W1	W2	W3		W1	W2	W3
개선_기대	2	4		부진_분기			6
개선_예상	1	5		부진_실적			6
개선_전망	1	5		부진_지속			6
개선_지속		6		부합_분기		6	
개선_추세		6		부합_실적		6	
견조_실적	5	1		분기_실적		6	
기다_리		1	5	분기_실적_부진		1	5
기대_부합		6		불확실성_해소	1	5	
기대치_부합		6		성장_모멘텀	2	4	
기대치_상회	6			성장세_지속	5	1	
기대치_하회			6	성주_목	1	5	
단기_실적	2	4		소폭_하회		1	5
대감_유효	1	5		수익성_개선	4	2	
대치_부합	2	4		수혜_주	1	5	
대치_상회	5	1		시장기_대치_부합		6	
매수_기회	2	4		시장기_대치_상회	6		
무난_실적		6		예상치_부합		6	
실적_개선	4	2		장기_성장	6		
실적_개선_지속	6			주주_환원	6		
실적_달성		6		지속_전망	2	4	
실적_리뷰		6		최대_수혜_주	6		
실적_모멘텀		6		컨센서스_부합	1	5	
실적_발표	1	5		컨센서스_상회	5	1	
실적_부진			6	컨센서스_하회		1	5
실적_예상		6		컨퍼런스콜요약	1	5	
실적_전망		6		탐방_노트	1	5	
실적_턴어라운드	2	4		필요시_점		1	5
실적_호조		6		하반기_실적_개선	6		
실적_회복		6		회복기_대	1	4	1
양호_실적		6		흑자_전환	1	5	
예상_부합	1	5		예상_수준		6	

〈표 4〉 강한 매수를 의미하는 제목 말뭉치, W1

단어 조합	부가설명	변수
견조_실적		강한 매수를 의미하는 제목 말뭉치, W1
기대치_상회		
성장세_지속		
수익성_개선		
시장기_대치_상회		
실적_개선_지속		
장기_성장		
주주_환원		
최대_수혜_주		
컨센서스_상회		
하반기_실적_개선		

〈표 5〉 매수를 의미하는 제목 말뭉치, W2

단어 조합	부가설명	변수
개선_기대		
개선_예상		
개선_전망		
개선_추세		
기대_부합		
기대치_부합		
단기_실적		
대감_유효	기대감 유효	
대치_부합	기대치 부합	
매수_기회		
무난_실적		
부합_분기		
부합_실적		
분기_실적		
불확실성_해소		
성장_모멘텀		
성주_목	가능성에 주목	
수혜_주		
시장기_대치		
시장기_대치_부합		
실적_달성		
실적_리뷰		
실적_모멘텀		
실적_발표		
실적_예상		
실적_전망		
실적_턴어라운드		
실적_호조		
실적_회복		
양호_실적		
예상_부합		
예상_수준		
예상치_부합		
지속_전망		
진행_중		
컨센서스_부합		
컨퍼런스콜요약		
탐방_노트		
하반기_실적		
회복기_대		
흑자_전환		

매수를
의미하는
제목 말뭉치, W2

〈표 6〉 매도를 의미하는 제목 말뭉치, W3

단어조합	부가설명	변수
기다_리	기다리며	매도를 의미하는 제목 말뭉치, W3
기대치_하회		
부진_분기		
부진_실적		
부진_지속		
분기_실적_부진		
소폭_하회		
실적_부진		
컨센서스_하회		
필요시_점	필요한 시점	

마지막 단계로는 세가지 말뭉치에 있는 단어 조합을 분석 데이터로 사용하기 위해 구조화된 정형데이터로 바꾸어야 한다. 본 연구에서는 세 가지 말뭉치가 각각의 제목에 존재하는지를 확인하기 위해 이항변수로 변환하였다. 이를 위해 R의 ‘tm’ 패키지를 이용하여 ‘용어 문서 행렬(Term Document Matrix)’을 만들어 제목에 말뭉치가 존재하는지 확인하였다.

3.3 재무분석가의 이익예측 특성 측정

본 연구에서는 선행 연구들을 바탕으로 연구 모형을 도출하였다. 최종적인 모형은 아래 식 (1), 식 (2)와 같다.

$$\begin{aligned}
 Error = & \alpha + \beta_1 W1 + \beta_2 W2 + \beta_3 W3 + \beta_4 LOSS \quad (1) \\
 & + \beta_5 ELV + \beta_6 BIG4 + \beta_7 SIZE + \beta_8 DIFF \\
 & + \beta_9 FOL + \beta_{10} EPS + \beta_{11} LEV + \beta_{12} ROA \\
 & + \epsilon
 \end{aligned}$$

$$\begin{aligned}
 TP_ANY = & \alpha + \beta_1 W1 + \beta_2 W2 + \beta_3 W3 \quad (2) \\
 & + \beta_4 LOSS + \beta_5 ELV + \beta_6 BIG4 + \beta_7 SIZE \\
 & + \beta_8 DIFF + \beta_9 FOL + \beta_{10} EPS + \beta_{11} LEV \\
 & + \beta_{12} ROA + \epsilon
 \end{aligned}$$

관심변수 외에 재무분석가의 이익예측 특성에

영향을 끼칠 수 있는 변수들을 선행연구를 참조하여 통제변수에 포함하였다. 통제변수로는 손실기업여부(LOSS), 이익수준(ELV), 감사품질(BIG4), 기업규모(SIZE), 예측기간차이(DIFF), 재무분석가 수(FOL), 주당순이익(EPS), 부채비율(LEV), 총자산 이익률(ROA)을 포함하였다.

손실기업여부(LOSS)는 해당 기업의 당해 당기순이익의 손실여부를 나타내는 변수이다. 당해 손실이 발생하는 경우, 해당 기업의 위험은 다른 기업보다 상대적으로 높고 경영자는 이익조정 유인에 노출될 가능성이 높아지게 된다 [10]. 따라서 손실기업여부는 이익예측 오차가 증가할 것이다. 이익수준(ELV)은 주당순이익을 기초주가로 나눈 변수로 전규안, 박종일[9], Eames and Glover[17]의 연구에서 이익수준이 높을수록 재무분석가의 이익을 낙관적으로 예측하여 주당순이익(EPS)변수와 함께 포함하였다. 감사인은 해당 기업의 재무제표 감사를 담당한 사람이며 감사인의 감사품질이 높아질수록 이익예측치의 정확성이 높아진다[14]. 따라서 감사품질(BIG4)의 측정은 대형 회계법인(삼정, 한영, 삼일, 안진)인지의 여부로 측정하였다. 기업규모(SIZE)는 해당 기업 자산의 자연로그 값으로 측정하였다. 기업규모는 기업을 나타내는 중요한 변수로서 기업규모가 큰 기업일수록 공시 정

보가 많으며[22], 더 많은 양의 공시자료는 기업의 미래 이익을 예측하는데 도움을 준다[4]. 기업규모가 큰 기업일수록 분석에 필요한 정보가 많기 때문에 재무분석가의 이익예측 오차는 감소한다[4, 16].

예측기간차이(DIFF)는 이익예측치 보고서 발행일과 실제 공시일간의 차이를 측정한다. 예측기간이 길어질수록 이익을 예측하기 위한 정보의 양이 부족하기 때문에 더 낙관적인 예측을 수행할 가능성이 있고, 이로 인하여 예측오차는 더 커지게 된다[5, 10].

재무분석가 수(FOL)는 기간 내 해당 기업에 대하여 예측치를 발표한 재무분석가 수를 나타내는 변수이다. 해당 기업에 대해 재무분석가의 수가 증가할수록 예측 오차가 줄어든다[22]. 부채비율(LEV)은 분석기업의 총부채를 총자산으로 나눈 값으로, 부채비율이 높아질수록 재무분석가의 예측오차가 커지게 된다[6, 18]. 총자산이익률(ROA)은 분석기업의 당기순이익을 자산총액으로 나눈 값으로 기업의 영업성과로 측정하였다. 총자산 이익률은 재무분석가의 이익예측오차와 음(-)의 관계를 갖는다[17].

IV. 연구 결과

4.1 기술 통계량

<표 7>은 기술통계량을 나타내는 것으로 총 표본 수는 10,318개이다. 종속변수인 이익예측치의 정확도(Error)와 이익예측의 달성(TP_ANY)의 경우 평균이 각각 0.463, 0.359, 중위수는 각각 0.328, 0.000으로 나타났다. 이익예측치의 정확도(Error)의 평균은 재무분석가의 이익예측치 정확도가 추정일 주가의 약 46%의 오차를 보인다는 것을 의미한다. 이익예측의 달성(TP_ANY)은 평균 약 36%를 차지하는 것으로 나타났다. 예측기간차이(DIFF)는 평균 약 234일, 중위수 약 243일로 나타났으며, 공시발표일 8개월 이전에 투자전

략 보고서가 많이 게시됨을 알 수 있다. 재무분석가 수(FOL)은 평균 약 127명, 중위수 112명, 최대값 389명으로 회계연도 1년 동안 상장되어 있는 한 기업에 대해 평균 약 127명의 재무분석가가 투자전략 보고서를 발행한다. 감사품질(BIG4)은 평균 약 97%로 나타나 대상기업의 대부분이 대형 감사법인으로부터 감사를 받는 것으로 나타났다. 그 외 손실기업여부(LOSS)는 평균 약 18%로 나타나 당기순이익이 (-)인 기업은 상장법인 중 약 18%를 차지함을 의미한다.

<표 7> 기술 통계량(N = 10,318)

Variable	Mean	Median	Std
Error	0.463	0.328	1.044
TP_ANY	0.359	0.000	0.480
DIFF	234.325	243.000	77.848
FOL	126.844	112.000	82.707
BIG4	0.973	1.000	0.162
EPS	11.190	11.127	0.209
ELV	0.032	0.042	0.088
LEV	0.425	0.428	0.199
LOSS	0.177	0.000	0.382
ROA	0.000	0.000	0.000
SIZE	29.340	29.336	1.546
w1	0.026	0.000	0.160
w2	0.072	0.000	0.258
w3	0.015	0.000	0.122

4.2 가설 검증결과

<표 8>은 제목으로부터 추출된 투자의견(W1, W2, W3)이 재무분석가의 이익예측 정확도(Error)에 미치는 영향과 관련된 가설 1에 대한 다중회귀분석 결과를 정리한 것이다. 분석 결과에 의하면, 전체 모형에 대한 설명력은 약 5%이며 F값 49.289, 유의확률은 $P < 0.001$ 로 유의한 결과를 보여주고 있다. 강한 매수 투자의견(W1)이 재무분석가의 이익예측 정확도(Error)에 미치는 영향을 나타내는 회귀계수 값은 -0.139로 5% 유의수준 미만(t 값 = -2.155)에서 통계적으로 유의한 결과를 보고하므로, 이는 강한 매수 투자의

견(W1)이 재무분석가의 이익예측 정확도(Error)와 음(-)의 관계가 있기 때문에 투자전략 보고서에 강한 매수 투자 의견(W1)이 있을수록 예측오차가 줄어들어 재무분석가의 이익예측 정확도(Error)는 높아짐을 의미한다. 강한 매도 투자의견(W2)이 재무분석가의 이익예측 정확도(Error)에 미치는 영향을 나타내는 회귀계수 값은 -0.174로 5% 유의수준 미만(t 값 = -2.108)에서 통계적으로 유의한 결과를 보고하므로, 이는 매도 투자의견(W3)이 재무분석가의 이익예측 정확도(Error)와 음(-)의 관계가 있기 때문에 투자전략 보고서에 매도 투자의견(W3)이 있을수록 예측오차가 줄어들어 재무분석가의 이익예측 정확도(Error)는 높아짐을 의미한다.

〈표 9〉는 제목으로부터 추출된 투자의견(W1, W2, W3)이 재무분석가의 이익예측 달성(TP_ANY)에 미치는 영향과 관련된 가설 2에 대한 다중회귀 분석 결과를 정리한 것이다. 분석 결과에 의하면,

모형의 적합도 검증에서는 모형에 대한 설명력은 (Nagelkerke R^2) 9%로 나왔다. 로지스틱 회귀분석 결과를 살펴보면, 독립변수의 회귀계수에 대한 통계적 유의성을 검증하는 값인 Wald 통계량과 유의확률 값을 통계적 유의수준 5%와 비교해보면, 로지스틱 회귀모형에 포함시킬 수 있는 변수로는 재무분석가 수(FOL), 주당순이익(EPS), 이익수준(ELV), 부채비율(LEV), 손실기업여부(LOSS), 총자산 이익률(ROA), 기업규모(SIZE), 매도 투자의견(W3)이 통계적으로 유의한 것으로 나타났다. 결과에 따른 해석은 다음과 같다. 재무분석가 수(FOL)의 변수 값들이 1단위 증가하면, 재무분석가의 이익예측 달성(TP_ANY)가 일어날 확률이 일어나지 않을 확률보다 1.002배 증가함을 의미한다. 매도 투자의견(W3) 변수 값들이 1단위 증가하면, 재무분석가의 이익예측 달성(TP_ANY)가 일어날 확률이 일어나지 않을 확률보다 2.356배 증가함을 의미한다.

〈표 8〉 투자의견과 이익예측 정확도(Error) 간의 관계

$$Error = \alpha + \beta_1 W1 + \beta_2 W2 + \beta_3 W3 + \beta_4 LOSS + \beta_5 ELV + \beta_6 BIG4 + \beta_7 SIZE + \beta_8 DIFF + \beta_9 FOL + \beta_{10} EPS + \beta_{11} LEV + \beta_{12} ROA + \epsilon$$

Variable	Coef.	t-value
Intercept	-10.645	-14.060***
DIFF	0.000	-0.835
FOL	-0.001	-7.288***
BIG4	0.333	5.238***
EPS	1.115	16.581***
ELV	1.367	8.403***
LEV	-0.537	-9.085***
LOSS	0.439	11.438***
ROA	-801.452	-7.302***
SIZE	-0.045	-4.399***
W1	-0.139	-2.155*
W2	0.019	0.479
W3	-0.174	-2.108*
F-value		49.289***
adj. R^2		0.053
N		10,318

* p < 0.05, ** p < 0.01, *** p < 0.001.

〈표 9〉 투자의견과 이익예측 달성(TP_ANY) 간의 관계

$$TP_{ANY} = \alpha + \beta_1 W1 + \beta_2 W2 + \beta_3 W3 + \beta_4 LOSS + \beta_5 ELV + \beta_6 BIG4 + \beta_7 SIZE + \beta_8 DIFF + \beta_9 FOL + \beta_{10} EPS + \beta_{11} LEV + \beta_{12} ROA + \epsilon$$

Variable	B	S.E.	Wals	유의확률	Exp(B)
DIFF	-1.928	1.594	1.463	.226	.145
FOL	.000	.000	1.754	.185	1.000
BIG4	.002	.000	31.117	.000	1.002
EPS	-.058	.130	.201	.654	.944
ELV	.904	.143	40.134	.000	2.468
LEV	-4.338	.365	140.894	.000	.013
LOSS	1.668	.129	165.988	.000	5.304
ROA	-1.064	.085	156.015	.000	.345
SIZE	-2507.818	305.037	67.591	.000	0.000
W1	-.314	.022	202.817	.000	.730
W2	-.183	.139	1.738	.187	.833
W3	.058	.085	.461	.497	1.059
상수항	.857	.170	25.363	.000	2.356
<i>NagelkerkeR²</i>	0.090				
N	10,318				

V. 결 론

본 연구는 재무분석가의 이익예측에 있어서 투자전략 보고서의 제목이 미치는 영향이 있는지를 분석한 것이다. 이에 본 연구는 한국거래소에 상장되어 있는 기업 중 2014년 04월 01일부터 2015년 03월 31일까지의 조건에 맞는 투자전략 보고서 제목 10,318개 관측치를 이용하여 투자전략 보고서 제목으로부터 추출된 투자의견에 따라 재무분석가의 이익예측에 미치는 영향을 분석하였다.

기존의 투자의견은 대부분이 매수의견이며 매도는 희박하였다. 따라서 매수의 의견을 더 자세히 볼 필요가 있다고 판단하였고, 투자의견을 강한 매수의견, 매수의견, 매도의견 총 세가지로 나누어서 분석하였다. 첫째, 투자전략 보고서의 제목의 의미에 따라 재무분석가의 예측 정확도에 영향을 끼치는지에 대해서 분석하였으며, 둘째는 투자전략 보고서의 제목의 의미에 따라 재무분석

가의 예측 주가 달성도에 영향을 끼치는지에 대해서 분석하였다. 분석결과에 따르면, 투자전략 보고서의 제목으로부터 추출된 강한 매수의견과 매도의견이 재무분석가의 예측 정확도에 통계적으로 유의미한 영향을 미치며, 재무분석가의 예측 정확도와 음(-)의 관련성이 있음을 발견하였다. 이는 두 중요변수들이 커질수록 예측오차가 작아지는 것을 의미하므로 결론적으로 재무 분석가의 예측 정확도가 높아지는 것으로 해석할 수 있다. 그리고 투자전략 보고서의 제목으로부터 추출된 매도의견만이 재무분석가의 예측 달성도에 통계적으로 유의미한 영향을 미치며, 재무분석가의 예측 달성도와 양(+)의 관련성이 있는 것으로 나타났다. 이는 매도의견이 있을수록 재무분석가의 예측 달성도가 높아지는 것을 의미한다. 다른 중요변수는 통계적으로 유의미하지 않게 나타났다. 이러한 결과를 종합해볼 때, 재무분석가의 투자전략 보고서의 제목이 재무분석가의 이익 예측에 영향을 끼치는 것을 알 수 있다.

본 연구는 투자전략 보고서의 제목에 따른 재무분석가의 이익 예측에 대한 영향을 분석함으로써 다음과 같은 시사점을 나타낸다. 먼저 학술적 시사점으로써, 첫째, 새로운 연구방법론을 이용한 재무분석가의 예측에 관련된 최초의 연구로써, 투자전략 보고서의 제목이 재무분석가의 이익예측에 영향을 미치고 있음을 확인하였다. 따라서 이러한 연구방법론을 이용하여 추가적인 관련연구를 수행할 수 있는 귀중한 밑거름이 될 것이다. 그리고 추후 연구에서 투자의견을 더 세분화한 후의 분석이 이루어진다면 새로운 시사점을 제공할 것으로 판단된다. 둘째, 기존의 재무분석가의 예측모형에 전략보고서 제목 말뚝치라는 새로운 변수를 추가하였다. 그리고 사회적 시사점으로, 본 연구는 투자자에게 더 정확한 재무분석가의 투자전략 보고서를 판단하는 가이드를 제시하였다.

본 연구는 다음과 한계가 있기에, 향후에는 이를 고려한 분석이 이루어져야 한다. 첫째, 재무분석가의 예측에 관련한 측정용 2가지 방법으로 국한 되었다. 특히 재무분석가의 달성도(TP_ANY)는 너무 관대한 측정방법이라는 것을 부정할 수 없다. 둘째, 분석 기간을 1년 단위인 점이다. 재무분석가의 투자전략 보고서의 목표주가라는 것이 어느 시기의 목표주가라는 것이 명시되어 있지 않기 때문에, 연구자가 주관적으로 그 시기를 정해야 한다. 따라서 분석 기간을 3개월 단위, 6개월 단위 등으로 나누어서 분석을 할 필요가 있다. 셋째, 본 연구의 표본에 포함되기 위해서는 반드시 재무분석가의 투자전략 보고서가 발행된 기업이어야만 한다. 따라서 기업규모가 작은 기업의 경우에는 표본이 제외되기 때문에 본 연구 결과를 일반화하기에는 주의해야 할 필요가 있다. 넷째, 본 연구의 분석단위는 제목이다. 기술통계량에 의하면 한 기업을 평가한 재무분석가 수(FOL)는 최소 5명이며 최대 389명이며 중위수는 112명이다. 이는 대상 기업을 평가한

재무분석가가 많은 회사의 기업정보가 작은 회사의 기업정보를 왜곡할 수가 있다. 다섯째, 본 연구에서는 형태소 분석을 명사형으로만 하였다. 재무분석가의 투자전략 보고서 제목이 한 줄 밖에 안되기 때문에 명사형으로만 분석하는 것도 의미가 있다고 보였다. 추후에는 형용사, 감탄사 등을 분석을 연구가 필요하다. 마지막으로는 형태소 분석기의 성능으로 인하여 결과가 명확하게 추출되지 않을 가능성이 있다. 자연어 처리과정 중 불필요한 분석을 없애기 위해서 ‘은’, ‘는’, ‘이’, ‘가’를 불용어 처리하였기 때문에 ‘능성’이라는 단어로 존재한다. 즉, ‘가능성’으로 추출하지 못하게 된다. 추후에는 형태소 분석기의 성능 개선으로 한계점을 보완한 연구가 이루어 질 것으로 기대한다.

참 고 문 헌

- [1] 고봉찬, 김진우, “애널리스트 이익예측의 정확성과 추천종목의 수익성”, 증권학회, 제36권, 제6호, pp.1009-1047, 2007.
- [2] 김동순, 엄승섭, “국내외 애널리스트들의 투자의견 및 목표주가 변경이 주가에 미치는 영향력 분석”, 증권학회지, 제35권, 제2호, pp.75-108, 2006.
- [3] 김유나, “내용분석을 통한 전원생활 및 전원주택의 패러다임 변화에 관한 연구”, 석사학위논문, 건국대학교대학원건축공학과, 2013.
- [4] 김지홍, 신현한, 장진호, “재무분석가수의결정요인: IBES 자료를 이용한 요인분석”, 회계학연구, pp.31-54, 2003.
- [5] 안윤영, 신현한, 장진호, “재무분석가의 명성과 성과간의 관계 및 명성의 결정요인”, 회계와 감사연구, 제43권, pp.375-402, 2006a.
- [6] 안윤영, 유영태, 조영준, 신현한, 장진호, “재무분석가의 특성이 이익예측 정확성에 미치

- 는 영향”, 회계학연구, 제31권, 제4호, pp.1-24, 2006b.
- [7] 이경주, “법원판례 분석을 통한 공동주택 하자 판결 시스템 개선에 관한 연구”, 박사학위 논문, 한양대학교건축공학과, 2014.
- [8] 이만원, 심현호, 장진호, “애널리스트의 이해 상충에 관한 연구”, 회계학연구, 제30권, 제3호, pp.173-194, 2005.
- [9] 전규안, 박종일, “세무상 이월결손금과 이익 조정에 관한 연구”, 한국회계학회, 하계학술대회발표논문집, 2005.
- [10] 정석우, “재무분석가의 분석기업결정과 예측 특성에 영향을 미치는 요인”, 회계학연구, 제 28권, pp.61-84, 2003.
- [11] An, J., W. Jung, and H.W. Kim, *A Green Flag over Mobile Industry Start-Ups: Human Capital and Past Investors as Investment Signals*, 2015.
- [12] Asquith, P., M.B. Mikhail, and A. Au., “Information Content of Equity Analyst Reports”, *Journal of Financial Economics*, Vol.75, No.1, pp.245-282, 2005.
- [13] Barber, B., R. Lehavy, M. McNichols, and B. Trueman, “Can Investors from the Prophets? Security Analyst Recommendation and Stock Returns”, *Journal of Finance*, Vol.56, pp.531-563, 2001.
- [14] Behn, B.K., J.-H. Choi, and T. Kang, “Audit Quality and Properties of Analyst Earnings Forecasts”, *The Accounting Review*, Vol.83, No.2, pp.327-349, 2008.
- [15] Bradshaw, M.T. and L.D. Brown., “Do Sell-Side Analysts Exhibit Differential Target Price Forecasting Ability? Working Paper”, *Harvard Business School*, MA, pp.1-44, 2006.
- [16] Brown, L., P. Griffin, R. Hageman, and M. Zmijewski, “Security analysts superiority relative to univariate time-series models in forecasting quarterly earnings”, *Journal of Accounting and Economics*, pp.61-87, 1987.
- [17] Eames, M.J. and S.M. Glover, *Earnings Predictability and the Directing of Analysts*, 2003.
- [18] Eddy, E. and B. Seifer, *An Estimation of Hypotheses Concerning Earnings Forecasting*, 1992.
- [19] Hearst, M.A., *Untangling Text Data Mining*, in Proceedings of the 37th ACL., 1999.
- [20] Hong, H. and J. Kubik, “Analyzing the Analysts: Career Concerns and Biased Earnings Forecasts”, *Journal of Finance*, Vol.58, pp.313-351, 2003.
- [21] Jagadeesh, N., J. Kim, S. Krische, and C. Lee., “Analyzing the Analysts: When Do Recommendation Add Value?”, *Journal of Finance*, Vol.59, pp.1083-1124, 2004.
- [22] Lang, M.H. and R.J. Lundholm, “Corporate disclosure policy and analyst Behavior”, *The Accounting Review*, Vol.71, pp.467-492, 1996.
- [23] Mooney, R.J. and R. Bunescu, “Mining Knowledge from Text using Information Extraction”, *ACM SIGKDD Explorations*, Vol.7, No.1, pp.3-10, 2006.
- [24] Sebastiani, F., “Machine Learning in Automated Text Categorization”, *ACM Computing Surveys*, Vol.34, No.1, pp.1-47, 2002.
- [25] Sinha, P., L.D. Brown, and S. Das, “A Re-Examination of Financial Analysts Differential Earnings Forecast Accuracy”, *Contemporary Accounting Research*, Vol.14, pp.1-42, 1997.
- [26] Spence, M., “Job Market Signaling”, *The Quarterly Journal of Economics*, Vol.87, No.3, pp.355-374, 1973.
- [27] Stanvrianou, A., P. Andritsos, and N. Nicoloyannis, “Overview and Semantic Issues of Text Mining”, *ACM SIGMOD Record*, Vol.36, No.3, pp.23-34, 2007.
- [28] Stickel, S.E., “Reputation and Performance Among

Security Analysts”, *Journal of Finance*, Vol.47, pp.1811-1836, 1992.

[29] Witten, I.H., *Text Mining, Practical Handbook of Internet Computing*, CRC Press, 2004.

[30] Womack, K.L., “Do Brokerage Analysts’ Recommendations Have Investment Value?”, *Journal of Finance*, Vol.51, pp.137-167, 1996.

저자 소개



장 준 규(Joon-Kyu Jang)

- 2014년 : 세종대학교 경영학과 (경영학 학사)
- 2016년 : 연세대학교 정보대학원(정보시스템학 석사)
- 관심분야 : Bigdata, Machine learning, Open collaboration



이 규 현(Kyu Hyun Lee)

- 2015년 : 연세대학교 정보통계학과 (이학사)
- 2015년~현재 : 연세대학교 정보대학원 (석사과정)
- 관심분야 : Statistical Business Analytics, Big Data Analytics, Social Media Mining



이 준 기(Zoonky Lee)

- 1985년 : 서울대학교 전산통계학과 학사
- 1991년 : 카네기멜론대학 사회심리학과 석사
- 1999년 : 남가주 대학교 경영정보학과 박사
- 2004년~현재 : 연세대학교 정보대학원 교수
- 관심분야 : Web2.0, E-Transformation, Dynamic pricing, KM, Open Innovation