

감성분석을 활용한 컨테이너 시장 예측[†]

Forecasting of Container shipping market using sentiment analysis

최건우*·윤희성**·김동원***

Choi, Gunwoo · Yoon, Heesung · Kim, Dongwon

목 차

- I. 서 론
- II. 감성분석방법 및 선행연구
- III. 감성분석을 활용한 컨테이너 시장 예측 결과
- IV. 결론

〈Abstract〉

This study attempted to analyze the container shipping market using sentiment analysis. Text data captured from Lloyd's List from January to December 2018 was used as input data. Three types of sentiment dictionaries, i.e. NRC, Bing and AFINN, were used to measure the sentiment of the text and the results were compared to the container spot market. The results of Bing and NRC correlated more than 60% with the direction of spot market movement, while AFINN generated 51% correlation. This research is meaningful in demonstrating the applicability of sentiment analysis to short-term forecasts in the freight market. A scientific approach to 'market sentiment' is expected to enhance the performance of shipping companies.

Key words: Sentiment analysis, Unstructured data, Container shipping

† 본연구는 한국해양수산개발원의 연구개발직립금사업인「감성분석을 활용한 해운시황 분석 연구」의 일부 내용을 발췌하여 재작성함

* 제1저자, 한국해양수산개발원 전문연구원, 경제학 박사.

** 교신저자, 한국해양수산개발원 부연구위원, 경영학 박사.

*** 제2저자, 한국해양대학교

I. 서 론

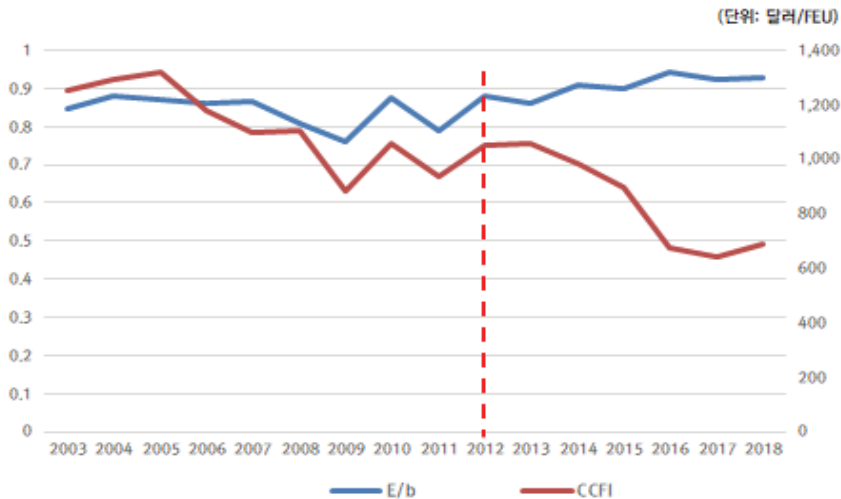
1. 연구배경 및 목적

2009년 세계 금융위기 이후 해운시장의 불확실성과 변동성은 더욱더 급증하였다. 2003년부터 발표되고 있는 중국발 컨테이너 운임지수(CCFI: China Containerized Freight Index)를 보면 2009년 이전의 주간단위 최대 변동폭은 2% 대에 불과하였으나 이후 3~6%대로 확대되는 것으로 나타났다. 또한 운임지수 최저치와 최대치간의 폭도 커지는 것으로 나타났다. 이렇듯 운임의 변동성이 심화되면서 선사가 이를 예측하고 대응하는 역량이 기업의 경영성과에 직접적으로 영향을 미치게 되었다. 문제는 경제·사회구조가 복잡화되면서 다양한 요인이 해운시장에 영향을 미치게 되어 선사가 이를 파악하는데 상당한 어려움을 겪고 있다는 점이다.

컨테이너 시장 예측은 대부분 시장의 수요와 공급에 관련된 시계열자료를 이용하여 계량적인 방법론을 활용하거나 전문가의 정성적인 예측을 통해 이루어진다. 최근에는 신경망분석과 같은 인공지능(AI: Artificial Intelligence)을 활용한 예측도 시도되고 있지만 분석시 이용되는 시계열자료가 가지고 있는 몇 가지 단점이 있다. 첫째, 자료의 주기에 따른 예측시점과 자료의 불일치이다. 계량적 또는 정량적 분석에 활용되는 자료는 대부분 시계열자료로 현재시점의 자료가 아닌 최소 1~2개월 이전의 자료를 활용한다. 이에 따라 예측시점이 아닌 1~2개월 이전의 시점의 상황에서 미래를 예측할 수밖에 없다. 둘째, 실시간으로 자료 수집이 불가능하여 외부충격에 대한 운임 예측이 어렵다. 올 초 브라질 댐 붕괴 사고로 인해 벌크시황이 최저 수준을 기록하였으나 이에 대한 계량적인 분석은 불가능하다. 하지만 기사, SNS와 같이 비정형자료를 이용하면 계량자료가 가지는 한계에서 벗어날 수 있다. 예를 들어 브라질 댐 붕괴 이후 발생한 관련 기사들에 대해 부정적 또는 긍정적 어휘의 개수만 분석하더라도 부정적 어휘의 비중이 크게 높았다.

컨테이너 운임 시장은 공급과잉으로 인해 수요와 공급 이외의 요소가 영향을 미치게 되었다. 실제로 아시아-북미 컨테이너 운임과 적재율과의 관계를 살펴보면 2012년 이후 컨테이너 운임은 적재율과는 전혀 무관한 움직임이 나타나는 것을 알 수 있다. 이는 컨테이너 시장에 충격이 있을 경우 시장 참여자들이 받아들이는 정도가 다르며 이는 시장 심리(sentiment)로 반영된다. 따라서 이러한 시장심리를 측정하고 상기에서 언급한 기존 자료의 한계를 보완하기 위해 비정형 자료를 이용한 해운시황 전망 가능성을 검토하는 것이 연구의 목적이다.

■ 그림-1. 컨테이너 적재율과 운임 ■



자료: Drewry, Container forecaster 2003~2019, 상해해운거래소

2. 연구의 내용과 방법

연구의 내용과 방법은 다음과 같다. 먼저 II장에서 감성분석방법론에 대한 설명과 관련된 선행연구를 검토한다. III장에서는 감성분석프로세스와 결과를 기술하고 컨테이너 시황 간의 관계 도출한다. 아울러 분석 자료의 작성방법과

분석결과를 설명하게 된다. IV장에서 연구결과 및 한계에 대해 기술한다.

II. 감성분석방법 및 선행연구

1. 감성분석의 개념

감성분석(sentiment Analysis)은 비정형 데이터를 분석하는 도구로서 글을 작성하거나 말을 하는 사람들의 태도, 성향을 분석하여 특정 주제에 대한 감성이 무엇인지 분류하는 것이다.(김영민, 2016)¹⁾ 제품 사용후기, 영화평 등의 온라인 리뷰나 트위터나 페이스북 같은 SNS를 통해 이슈나 의견(opinion)으로부터 텍스트의 감성이 긍정인지 혹은 부정인지 혹은 중립인지를 알아내는 것이다. 서비스 또는 상품사용자의 의견을 분석한다고 해서 일명 오피니언 마이닝(opinion mining)이라고 불리기도 한다. 감성분석은 화자나 글쓴이가 표현한 말이나 글에서 감성어휘를 추출하며, 감성사전을 활용하여 긍정과 부정 또는 슬픔, 기쁨, 분노 등 다양한 감성을 추출하여 감정상태를 평가하는 분석기법으로 텍스트 마이닝(text mining)의 한 종류이다.

분석에 사용되는 데이터 형태는 정형 데이터와 비정형 데이터로 분류되고 있는데 이번 연구에서 이용되는 감성분석에 사용되는 데이터는 비정형 데이터이다. 정형 데이터는 숫자 데이터뿐만 아니라 제품 주문 시에 입력하는 주소, 연락처, 결제 정보와 같이 매번 사용할 때 마다 데이터베이스 내의 고정된 필드에 저장되는 데이터를 말한다. 반대로 비정형 데이터는 고정된 필드에 저장되지 않은 데이터를 일컬으며 트위터(Tweeter)나 페이스북(Facebook)과 같은 SNS에 업로드 된 댓글, 동영상, 사진, 메신저 상의 대화문, 이메일(e-mail), 그리고 전화 통화 내용 등이 이에 해당된다.

1) 김인경, 김혜민, 임병환, 이기광(2016), 「감성분석 결과와 사용자 만족도와의 관계-기상청 사례를 중심으로-」, 한국콘텐츠학회 논문지, Vol.16, No. 10, p.3

2. 감성분석의 종류 및 분석절차

감성분석은 분석 대상인 텍스트의 범위에 따라 문서단위(단위(Document-level), 문장단위(Sentence-level), 속성단위(Aspect-level)의 분석으로 분류할 수 있다.(명은진, 2018)²⁾ 문서-문장-속성단위로 갈수록 세분화하여 감성을 분석하는 방법으로 문서단위 분석은 한 개의 문서 전체에서 한 개의 감성을 가지고 있다고 가정하여 분석한다.

감성분석을 분석을 위해 사용한 방법론에 따라 분류할 수도 있는데 크게 두 가지 방법론이 있다. 한 가지는 기계학습 기법 (machine learning approach)이고, 다른 하나는 감성사전을 기반으로 한 사전 기반 기법 (lexicon-based approach)이 있다. 기계학습 기법은 데이터 일반적으로 대량의 문서에 사용되는데 토큰화(tokenization) 등을 거쳐 데이터는 전처리되며 이를 ‘학습’을 통해 긍정과 부정과 같은 감성분류(class)를 실시하는 방법이다. 일반적으로 기계학습 분류기로 나이브 베이즈 방법과 서포트벡터머신(SVM: Support Vector Machine)를 활용한다. 사전 기반 방법은 연구자가 감성사전을 구축하고 분석하고자 하는 대상에 대해 감성사전과 매칭을 통해 최종적으로 감성을 분류하는 방식이다.

3. 선행연구 검토

감성 분석에 대한 연구는 1990년대 후반부터 시작되어 2000년대에 이르기까지 점차적으로 증대한 연구 분야로 자리매김하고 있다.³⁾ 주로 제품 구매 리뷰나 영화에 대한 리뷰에 대한 의견을 긍정적이거나 부정적인 감성으로 분류하는 것에 초점을 맞추어 연구해 오고 있다. 이러한 분류 방법은 감성 분석의 연구 영역에 더 깊은 이해와 제안을 제공하는 데 중요한 역할을 하고 있다.⁴⁾ 이

2) 명은진(2018), 「감성사전 기반 Word2vec 자질을 이용한 감성 분류 시스템」, 석사학위논문, p.4

3) Thelwall, M., Buckley, K., & Paltoglou, G., 2008. Sentiment in Twitter events. Journal of the American Society for Information Science and Technology, 62, pp.406-418.

러한 연구 문헌들이 중점적으로 다루고 있는 양상은 부정적이고 긍정적인 고객의 리뷰가 제품과 서비스의 장점과 단점을 보여주고 있는 것으로 추정하고 있다는 것이다. 동 분석방법은 주로 상품 및 서비스에 대한 평가요소를 측정하는데 사용되었지만 최근에는 금융분야를 중심으로 전망 및 예측분야에서도 활용되고 있으며 현재까지 해운분야에 적용된 사례는 없다.

Papapostolou 외(2016)⁵⁾는 해운의 주요 지표를 활용하여 감성지수(sentiment index)를 생성하여 7개 국가(브라질, 러시아, 인도, 중국 등) 주식과 비교 분석했다. 감성지수의 생성은 컨테이너, 벌크, 탱커 분야로 분류했으며 확실히 등가수익(CER: Certainty Equivalent Return)측면에서 볼 때 탱커 감성지수의 통계적 유의도가 가장 높은 것으로 분석되었다. 동 연구는 해운 지표를 이용해 감성지수를 개발하여 실물 경제를 분석한 연구로 시장 참여자의 감성을 활용한 사례로 의의가 있다.

Yuan Song, Hongwei Wang and Maoran Zhu(2018)⁶⁾는 일반적인 재무분석과는 다르게 기업의 사회적 책임(CSR: Corporate Social Responsibility)에 대한 투자자의 감성분석을 통한 의사결정에 대해 연구하였다. 투자자의 사회적 책임과 같이 비재무적분야에 대해 서포트벡터머신(SVM: Support Vector Machine)을 활용한 감성분석을 실시한 결과 분류 정확도는 86.0% 이상으로 나타났다.

Loughran과 McDonald(2011)⁷⁾는 감성적 표현이 영역마다 다르다는 것을 발견했다. 그들은 구체적으로 금융 분야에서 이미 구축된 감성사전의 지표가 이전의 모든 분야에 일반적으로 적용될 수 있다고 여겨졌던 것이 잘못된 것이

4) Wilson, T., Wiebe, J., & Hoffmann, P., 2009. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational linguistics*, 35, pp. 399-433.

5) Nikos C. Papapostolou, Panos K. Pouliasis, Nikos K. Nomikos, Ioannis Kyriakou(2016). Shipping Investor Sentiment and International Stock Return Predictability, *Transportation Research Part E*, 96, pp. 81-94.

6) Yuan Song, Hongwei Wang and Maoran Zhu(2018), 「Sustainable strategy for corporate governance based on the sentiment analysis of financial reports with CSR」, *Financial Innovation*, Vol 4, no.2

7) Loughran, T., & McDonald, B., 2011. When is a liability not a liability? Textual analysis dictionaries, and 10-Ks. *Journal of Finance*, 66(1), 35-65.

며, 감성의 지표는 각 분야별로 다르다고 주장하였다. O'Leary(2011)⁸⁾는 일반적으로 긍정 및 부정 감성이 시장에서의 부정적인 행동을 설명하는 데 몇 가지 제한사항을 갖고 있으며, 분석의 품질을 향상시키기 위해 도메인별 어휘들을 설명할 것을 제안하였다.

김동영(2015)⁹⁾은 KOSPI 상장사 중 7개 기업의 기사와 SNS 자료를 감성분석과 기계학습을 통한 예측모델을 개발하였다. 동 연구에서는 긍정지수(Positive index)를 산정한 후 익일 주가변동 항목을 추가해 기계학습을 위한 데이터로 사용하였다. 분석 결과 80%가 넘는 정확도를 기록하였다.

박성화(2018)¹⁰⁾는 1991년 1월부터 2018년 3월까지 18,909건의 뉴스 빅데이터를 활용하여 우리나라 해운업의 신용위험을 측정하는 해운뉴스지수를 개발하였다. 동 지수를 작성하기 위해 Grammenos(2010) 신용평가 6C에 정의한 5개부문 26개 키워드가 포함된 언론기사의 빈도를 사용하였다. 동 지수는 해상 운송화물 수입과의 유의미한 관계가 있는 것으로 나타나 해운시장의 위험을 감지하고 예측하는데 기여할 수 있는 것으로 판단된다.

김인겸(2016)¹¹⁾ 외는 기상청에서 실시하고 있는 만족도 설문조사를 보완하기 위해 소셜네트워크(SNS)를 활용한 감성분석을 실시하였다. 자료는 2011년~2014년 트위터 자료를 활용하였으며 나이브 베이즈 방법을 통해 긍정, 부정 및 중립적인 감정으로 분류하였다. 이를 위해 추가사전을 만들어 정확도를 향상시켜 약 97%의 정확도를 기록하였다.

정근웅 외(2018)¹²⁾는 온라인 게임 판매에 전자구전 효과가 어떠한 요인들이 영향을 미치는지 분석하였다. 이를 위해 서포트벡터머신을 활용하여 감성지수를 산정하였다. 산정된 감성지수와 전자구전효과의 크기인 리뷰, 평점 등을

8) O'Leary, D.E., 2011. Blog mining-review and extensions: from each according to his opinion. Decision Support Systems, 51(4), 821-830.

9) 김동영, 2015, 「SNS와 뉴스기사의 감성분석과 기계학습을 이용한 주가예측 모형에 관한 연구」, 학위논문

10) 박성화, 2018, 「뉴스 빅데이터를 이용한 해운뉴스지수 개발」, 해양정책연구, Vol.33 no.1, pp. 281-302

11) 김인겸, 김혜민, 임병환, 이기광, 2016, 「감성분석 결과와 사용자 만족도와와의 관계」, 한국콘텐츠학회논문지, Vol.16, no. 10 pp. 393-402

12) 정근웅, 김종욱, 2018, 감성분석 기반의 게임소비자 온라인 구전효과, 한국디지털정책학회논문지, Vol.16. No.3, pp.145-156,

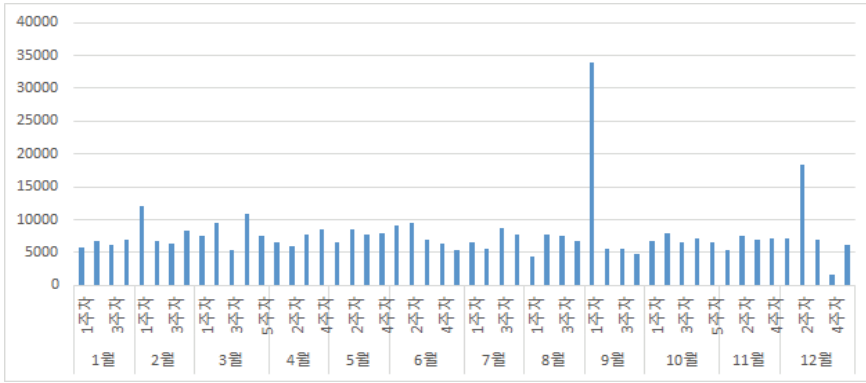
활용하여 회귀분석 하였다. 분석 결과 전자구전효과의 감성지수가 온라인 게임 판매에 영향을 미치는 것으로 나타났다.

Ⅲ. 감성분석을 활용한 컨테이너 시장 예측 결과

1. 자료

컨테이너 시황에 대한 자료 수집은 해운전문저널인 로이즈리스트(Lloyd's List)에서 자료를 수집한다. 국내 저널을 이용할 수도 있으나 국내저널이 해외 전문저널의 기사를 인용하는 비중이 높고 해운시장에서 국내시장의 영향도가 크지 않다는 점이 고려되었다. 분석 자료는 저널의 컨테이너 카테고리에 있는 모든 기사이며 수집기간은 2018년 1월 1일부터 2018년 12월 31일까지이며 주 단위로 자료를 수집하였다. 분석자료 중 컨테이너 운임과 관련된 자료를 필터링 하여 분석할 경우 연구 결과는 향상될 수 있으나 편향(bias)된 연구 결과일 수 있으므로 컨테이너 섹터에 있는 모든 기사에 대해 분석하였다. 이번 연구에서는 전처리되어 분석에 총 사용되는 토큰(token)은 총 404,332개이다. 감성분석결과와 비교분석하는 해운시황은 해운운임지수인 중국발 컨테이너 운임지수(CCFI)를 사용하였다. 일반적으로 컨테이너 운임을 대표하는 지수는 상해발 컨테이너 운임지수(SCFI)와 중국발 컨테이너 운임지수(CCFI)가 있다. 두 운임지수는 조사 패널 범위의 차이만 있을 뿐 상승과 하락은 거의 동일하게 움직이기 때문에 이번연구에서는 어느 지수를 사용하더라도 유사한 연구결과를 얻을 수 있다. 아울러 중국발 컨테이너 운임지수(CCFI)는 상해발 컨테이너 운임지수에 비해 변동성이 상대적으로 낮다.

■ 그림-2. 해운시황 분석에 사용되는 토큰(token) 추이 ■



2. 감성분석 절차

컨테이너 해운 시장에 대한 감성분석을 실시하기 위해서는 몇 가지 절차를 거치게 된다. 첫째, 데이터의 수집이다. 구체적으로 정기적으로 발행하는 해운 시장에 대한 비정형 자료를 수집하여 저장해야 한다.

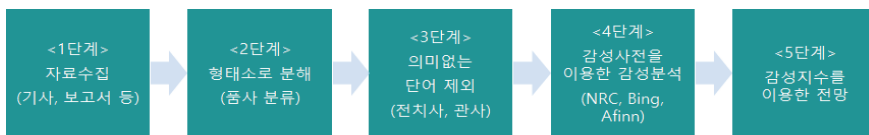
둘째, 수집된 자료를 뜻을 가진 최소단위인 형태소로 분리하고 이를 품사로 태깅(tagging)하여 정의한다. 또한 관사, 전치사 등 의미가 없는 어휘와 특수문자와 같은 기호는 필터링하여 제외한다. 이를 통해 필요한 정보만을 형태소로 분류하여 저장한다. 이러한 과정을 토큰화(tokenization)라 일컫는다.

셋째, 감성분석 단계로 어휘가 가지고 있는 긍정과 부정의 의미를 빈도수를 이용하여 측정한다. 이를 위해서 감성사전(Sentiment Dictionary)을 이용한다. 감성분석에 사용되는 툴(tool)인 R에서는 세 가지(NRC, Bing, Afinn) 종류의 감성사전을 제공한다. Bing은 총 6,778단어에 대해서 긍정(positive)과 부정(negative)로 구분한다. 긍정과 부정의 횟수를 통해 자료가 어떠한 감성을 가지고 있는지 판단한다. NRC는 총 13,891개 단어에 대해 긍정(positive)과 부정(negative) 뿐만 아니라 노여움(anger), 기대(anticipation), 혐오(disgust), 공포(fear), 즐거움(joy), 슬픔(sadness), 놀람(surprise), 신뢰(trust)의 총 8가지 감정

으로 구분한다. 예로 ‘abandon’이라는 단어는 부정적이면서 슬픔과 공포를 가지고 있는 단어로 분류한다. NRC 감성사전으로 이용하면 긍정과 부정뿐만 아니라 8가지 감정에 대한 빈도수를 확인할 수 있다는 강점을 가지고 있다. AFINN은 적용 어휘수가 2,000여개 불과하지만 긍정(positive)과 부정(negative)단어를 -5~5사이의 척도로 나타낸다. 단어가 가지고 있는 ‘세기’를 정량화하여 실세 수치로 볼 수 있게 해주는 감성사전이다. 이와 같은 방법은 단어에 따라 척도가 다르기 때문에 작성자의 감성을 척도로 표현해준다는 장점이 있지만 여러 사람이 작성한 기사에서는 객관적인 지수를 제공하기 어렵다는 단점을 가지고 있다. NRC와 Bing 사전에서는 부정과 긍정의 어휘에 대한 빈도수를 측정하여 계량화한다. AFINN은 적용 단어에 대해 긍정과 부정의 척도가 부여되기 때문에 이를 모두 집계하여 산정하였다.

마지막으로 분석된 감성지수와 해운시황과 비교·분석하여 감성지수의 이용 가능성을 검토한다. 해운시황과의 정확도 비교에서는 컨테이너 운임지수와 감성분석에서 도출된 감성(지수)의 상승과 하락의 일치도를 통해 판단하였다. 다시말해 특정 시점(T)의 부정적인 단어의 빈도수가 전주대비 증가하였다면 다음 주(T+1)시점의 컨테이너 운임이 하락할 경우 예측 정확도가 일치한 것으로 산정하였다.

■ 그림-3. 컨테이너 시장 감성분석 절차 ■



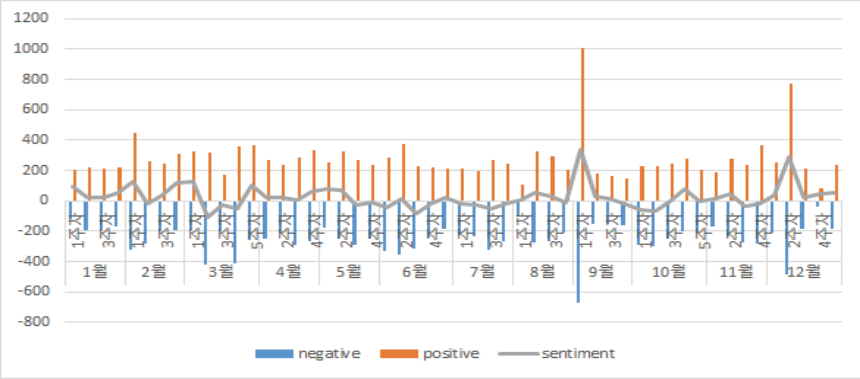
자료: 저자 작성

3. 감성분석 결과

감성사전을 이용하여 로이즈리스트(Lloyd's List)를 분석한 결과 긍정과 부정의 감성 변화는 유사하게 움직이는 것으로 나타났다. 이러한 감성의 변화는

기사의 양과 밀접하게 관련이 있는 것으로 나타나 총 감성토큰 중 부정과 긍정 또는 8가지 감성의 변화율을 컨테이너 운임지수와 비교하였다. 그 결과 긍정적인 단어의 빈도수가 부정적인 단어에 비해 높은 것으로 나타났는데 다른 해운 저널인 JOC, Tradewinds에서도 유사한 결과를 얻어 감성사전에서 긍정적인 어휘의 비중이 높은 것이 영향을 미친 것으로 추정된다.

■ 그림-4. ‘Bing’을 활용한 로이즈리스트 감성분석 결과 ■



주: 부정적인(negative)는 음수로 표시함

‘Bing’감성사전을 활용한 분석과 컨테이너 운임의 등락을 분석한 결과 부정적 단어와 컨테이너 운임 등락 간의 적중률은 약 63.3%인 것으로 분석됐다. 부정적인 단어의 예측이 긍정적인 단어에 비해 예측력이 높아졌으며 감성지수는 가장 낮은 결과를 기록하였다.

■ 표-2. 감성분석을 활용한 컨테이너 시장 예측 결과(Bing) ■

구분	부정적	긍정적	감성
컨테이너시황과 일치 횟수	31	30	19
정확도(%)	63.3	61.2	38.8

주: 운임 미발표(국경절, 춘절) 제외

‘NRC’ 감성사전을 활용한 분석에서는 혐오(61.2%), 놀라움(61.2)을 가지고 있는 단어와 컨테이너 운임 간의 관계가 가장 높은 것으로 나타났으며 즐거움(49.0%)의 적중률이 가장 낮은 것으로 나타났다.

■ 표-3. 감성분석을 활용한 컨테이너 시황 예측 결과(NRC) ■

구분	분노	기대	혐오	공포	즐거움	슬픔	놀라움	신뢰
컨테이너시황과 일치 횟수	26	28	30	26	24	28	30	27
정확도(%)	53.1	57.1	61.2	53.1	49.0	57.1	61.2	55.1

‘Afinn’ 감성사전을 활용한 분석에서는 정확도는 약 51.0%를 기록한 것으로 나타났다. ‘NRC’감성사전에 비해 분석 가능한 어휘가 15%에 불과하여 정밀한 분석에는 한계가 있는 것으로 판단된다.

■ 표-4. 감성분석을 활용한 컨테이너 시황 예측 결과(Afinn) ■

구분	감성지수
컨테이너시황과 일치 횟수	25
정확도(%)	51.0

IV. 요약 및 결론

본 연구에서는 불확실성과 변동성이 높은 해운시장을 분석하기 위해 현재까지 분석에 이용되지 않은 기사, 보고서와 같은 비정형 자료를 활용 가능성을 검토했다. 비정형자료를 분석하기 위해 텍스트 마이닝의 일종인 감성분석을 도입하였다. 일반적으로 감성분석은 제품, 서비스 평가분야에 활용되고 있으며 긍정과 부정과 같은 소비자의 의견을 청취하는데 사용되어 오피니언 분석

(opinion analysis)으로 불리지만 이번 연구에서는 소비자의 의견이 아닌 전문가들이 바라보는 해운시장에 대한 인식을 측정하는 도구로 활용하였다.

해운시장의 감성분석을 위해 해운 전문 저널인(Lloyd's List)를 활용하였으며 3가지('Bing', 'NRC', 'Afinn') 감성사전을 이용하여 긍정과 부정에 대한 어휘 뿐만 아니라 노여움(anger), 기대(anticipation), 혐오(disgust), 공포(fear), 즐거움(joy), 슬픔(sadness), 놀람(surprise), 신뢰(trust)와 같은 다양한 감성을 분석하여 컨테이너 운임시장 예측에 적용하였다.

감성분석 결과 'Bing'에서 분석된 부정적 단어의 빈도수가 컨테이너 해운시장과의 적합도가 가장 높은 것으로 나타났으며 대부분의 감성이 50~60%대의 정확도를 기록했다. 정확도만으로 볼 때 컨테이너 해운시장을 분석하기에는 부족하지만 이번 연구는 몇 가지 의의와 한계를 가지고 있다.

첫째, 실시간 분석을 통해 시장에 대한 인식을 측정하는 방안을 검토하였다. 현재 해운 시장을 분석하는 대부분의 분석 방법은 수요와 공급에 관련된 숫자 기반의 정형화된 자료를 이용한다. 회귀분석 기반의 연립방정식, 시계열 기반의 Box-Jenkins 방법론 등이 활용되고 있으나 활용되는 자료와 분석시점간의 시차가 존재한다. 다시 말해, 실제 분석하는 시점에서의 자료는 최소 1~2개월 이전 자료를 활용하여 예측과 자료 간에 일정기간 갭(gap)이 존재하는 문제가 항상 발생한다. 하지만 비정형 자료는 실시간으로 자료 수집 가능하여 기계학습이나 사전기반기법을 통해 즉시 분석이 가능하다는 장점이 있다.

둘째, 전문가의 직관을 계량화하였으며 전통적인 계량분석에서 벗어난 새로운 자료의 형태를 활용하였다. 해운 시장을 바라보는 인식을 정성적인 기준이 아닌 텍스트 마이닝의 한 종류인 감성분석을 활용하여 객관·정량화를 시도했다. 또한, 기존 해운시황에 이용되지 않은 텍스트자료의 활용 가능성을 검토하였다. 마지막으로 감성분석 결과가 새로운 변수로서의 기능을 할 수 있다. 감성분석을 통해 긍정, 부정, 다양한 감성에 대한 지수를 산출 할 수 있을 뿐만 아니라 기사의 양도 측정할 수 있다. 이들 변수는 해운시장을 예측하는 새로운 변수로서 활용 가능하다.

셋째, 상기에서 기술한 바와 같이 새로운 방법론을 통해 해운산업의 심리(sentiment) 측정하여 불확실성을 최소화할 수 있게 되었다. 이번 연구는 감성분석 방법 중 감성사전을 활용하는 방법론으로 시장의 감성(sentiment)를 분석하였지만 향후 발전적인 연구를 통해 개선이 될 경우 해운 시장 참여자에게 긍정적인 효과를 가져 올 것으로 기대된다.

이번 연구는 상기와 같이 몇 가지 연구 의의를 가지고 있지만 한계를 보완하기 위한 추가연구가 필요하다. 첫째, 분석 어휘에 대한 긍정과 부정의 속성과악에 대한 세부적인 연구가 필요하다. 예를 들어 ‘증가하다’(increasing, rising)와 같은 단어는 긍정의 의미로 분류된다. 하지만 해운시장에서 공급량이 늘어나는 경우 운임이 하락하여 오히려 부정적인 어휘에 가깝다. 이는 문맥 전체를 이해하여야 하는데 이번 연구에서는 감성사전을 활용한 사전기반 기법(lexicon-based approach)을 활용하였다. 향후 기계학습 기반의 감성분석이 진행되면 현재 측정하는 방법에 비해 나은 결과를 기대할 수 있다. 둘째, 기사와 해운시황과의 시차에 대한 연구도 진행되어야 한다. 이번 연구에서는 t시점의 기사를 활용하여 t+1시점을 예측하였지만 이를 조정하면 더 나은 결과를 얻을 수 있으므로 이에 대한 연구도 필요하다. 아울러 예측 시점에 대한 연구도 진행되어야 한다. 1년 동안 발행되는 기사를 수집해서 분석한 결과 대부분의 기사는 현재 시점에서의 시장의 이슈를 전달하였다. 하지만 중기, 장기 시장과 관련된 기사도 발행되므로 이번 연구의 한계는 명확하며 향후 머신러닝 기반으로 분석할 경우 보완될 것으로 생각된다. 셋째, 이번 연구는 텍스트 마이닝 중 기초적인 단계로 운임지수의 등락 예측을 목표로 하였지만 향후 시황에 대한 정교한 감성지수가 개발될 경우 레벨(level)에 대한 예측도 가능할 것이다. 마지막으로 다양한 비정형 자료의 접근성을 확보해야 한다. 자료의 범위가 넓어질수록 편차는 작아지게 되며 예측력 또한 상승할 것으로 기대되므로 해운 자료를 수집, 저장할 수 있는 저장고(storage) 체계마련 연구가 향후에 수행되어야 할 것이다.

투고일	2019.10.31
1차 심사일	2019.11.28
게재확정일	2019.12.09

■ ■ 참고문헌

1. 김동연, 2015, 「SNS와 뉴스기사의 감성분석과 기계학습을 이용한 주가예측 모형에 관한 연구」, 학위논문
2. 김인겸, 김혜민, 임병환, 이기광, 2016, 「감성분석 결과와 사용자 만족도와의 관계」, 한국콘텐츠학회논문지, Vol.16, no. 10 pp. 393-402
3. 명은진(2018), 「감성사전 기반 Word2vec 자질을 이용한 감성 분류 시스템」, 석사 학위논문
4. 박성화, 2018, 「뉴스 빅데이터를 이용한 해운뉴스지수 개발」, 해양정책연구, Vol.33 no.1, pp. 281-302
5. 정근웅, 김종욱, 2018, 감성분석 기반의 게임소비자 온라인 구전효과, 한국디지털 정책학회논문지, Vol.16. No.3, pp.145-156,
6. Loughran, T., & McDonald, B., 2011. When is a liability not a liability? Textual analysis dictionaries, and 10-Ks. *Journal of Finance*, 66(1), pp. 35-65.
7. Nikos C. Papapostolou, Panos K. Pouliasis,, Nikos K. Nomikos, Ioannis Kyriakou(2016). Shipping Investor Sentiment and International Stock Return Predictability, *Transportation Research Part E*, 96, pp. 81-94.
8. Thelwall, M., Buckley, K., & Paltoglou, G., 2008. Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology*, 62, pp. 406-418.
9. Wilson, T., Wiebe, J., & Hoffmann, P., 2009. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational linguistics*, 35, pp. 399-433.
10. Yuan Song, Hongwei Wang and Maoran Zhu(2018), 「Sustainable strategy for corporate governance based on the sentiment analysis of financial reports with CSR」, *Financial Innovation*, Vol 4, no.2 pp. 1-14