



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

物流學碩士 學位論文

빅데이터를 활용한 부산항 환적화물 물동량
이상치 탐지 분석

Detecting abnormal changes in transshipment
cargo throughput
in Busan Port using big-data



지도교수 신재영

1945

2018年 2月

韓國海洋大學校 海洋金融·物流大學院

海運港灣物流學科

박미혜

本 論文을 박미혜의 物流學碩士 學位論文으로 認准함.

위원장 남 기 찬 (인)

위 원 신 용 준 (인)

위 원 신 재 영 (인)

2017 년 12 월 27 일

한 국 해 양 대 학 교 해 양 금 용 물 류 대 학 원

< 요약 >

부산항의 환적화물은 최근 5년 동안 연평균 8.27%의 증가율을 나타내고 있다. 이는 국내 환적화물 물동량의 95%를 차지하나, 2016년 9월 한진해운사태를 비롯해 미국 대통령 트럼프의 보호무역주의, 중국 환율 조작국 지정 등의 경제 기조와 세계 교역량의 감소와 선대의 축소 등으로 인한 물량 감소는 부산항에 좋지 않은 영향을 주고 있다. 그러나 현재 선사에서 마케팅을 위해 활용하고 있는 자료로는 부산항에 신고된 수출입 환적 화물의 물량 통계뿐이다. 물론 원시 데이터는 있지만 공개되지 않고 있고, 월별 년별 집계된 데이터가 대부분이다.

그래서 부산항 전체 뿐만 아니라 다른 국가, 항만, 각 선사별로 부산항으로 환적 되는 물량의 변화가 어떤지, 특별히 나타나는 이상치는 없는지 분석하는 자료가 필요하다.

본 논문은 변화되고 있는 환적화물의 흐름을 빅데이터 기반으로 분석하고, 이를 통해 부산항이 데이터 기반 마케팅 관점에서 환적 화물의 부산항 글로벌 경쟁력 강화 전략을 수립하는데 도움을 주는 것을 목표로 한다. 분석을 위한 자료는 부산항만공사의 Port-MIS 자료 중 부산항을 경유하는 환적화물 자료 및 이를 활용한다. 또한 터미널에서 관리하고 있는 COARRI, CODECO 자료도 포함했다. 부산항 전체는 시계열분석 방법의 하나인 ARIMA 분석 방법을 사용하였고, 국가별, 포트별 선사별 분석에서는 이동평균법(MA)방법을 사용하여 신뢰수준 95%하에서 벗어나는 것을 이상치로 간주하였다. 또한 해당 이상치를 감지한 국가 및 항만에 대해서 상세하게 분석하여 어떤 국가에서 어떤 항만, 어떤 선사의 환적화물 물량이 이상이 있는지 그 원인은 무엇이며, 다른 경쟁 항만은 어떤 일이 일어나고 있는지 환적화물 물동량 패턴을 분석하여 항만공사, 항만청, 선사, 터미널 운영사 등 부산항을 이용하는 항만 주체들의 의사 결정에 도움을 주고자 한다.

< Abstract >

The transshipment cargo of Busan Port have shown an annual increase of 8.27% over the past five years. This is 95% of transshipment cargo in Korea. However, due to the Hanjin Transportation case in September 2016, also the protectionism from United States president Trump, including setting up a currency manipulator, China's economy, Due to reduction in fleet with a reduction in world trade and its volume reduction is a bad influence on the southern port of Busan. However, the only data that the shipping companies are using for marketing is just a rate of the volume of transshipment cargo reported in Busan Port. Of course, there are raw data but not released, and most of the data is aggregated monthly by year.

Therefore, it is necessary to provide data to analyze whether there is any abnormal change in the volume of transshipment to ! Busan by not only the entire port of Busan but also other countries, ports, and shipping companies.

The purpose of this paper is to analyze the changing flow of transshipment goods on a big data base and aims to help the Busan Port establish a global competitive advantage strategy for transshipment goods from a data-driven marketing perspective. Data for analysis are based on the transshipment cargo data via Busan Port, among the Port-MIS data of Busan Port Authority. It also included COARRI and CODECO data managed by the terminal. For entire Busan Port used ARIMA, one of the time-series analysis methods, and it was considered to be out of the range of 95% confidence level using the MA method in the analysis of shipper by country and port. It also analyzes the countries that detected such anomalies in detail and ports to determine which ports or ports are operating in the country, and determines the reasons for the failure of the port owner's transshipment and transportation freight terminal operations, and determines other competing ports.

< 목 차 >

Abstract	1
제1장 서론	8
1.1 연구의 배경과 필요성	9
1.2 연구의 방법	9
제2장 이론적 고찰	11
2.1 빅데이터의 개념과 특징	11
2.1.1 빅데이터의 정의	11
2.1.2 빅데이터 출현배경	12
2.1.3 빅데이터 특징	15
2.2 빅데이터 예측 분석 기법	18
2.2.1 빅데이터 분석 기법	18
2.2.2 시계열모형	20
2.3 해운 항만 물류 빅데이터 연구 사례	24
2.3.1 해운 항만 물류 빅데이터 연구 사례	26
2.3.2 물동량 변화, 예측 연구 사례	29
제3장 부산항 환적화물 물동량 현황 및 흐름 패턴 분석	35
3.1 부산항 환적화물 물동량 현황	35
3.2 부산항 환적화물 물동량 흐름 패턴 분석	36

제4장 부산항 환적화물 물동량 이상 감지	42
4.1. 현행 부산항 환적화물 물동량 이상 감지 방법	42
4.1.1. 현행 환적 화물 물동량 이상 변화 감지 방법	42
4.1.2. 시계열 모형을 이용한 환적화물 물동량 이상 감지 방법	43
4.2. 부산항 환적화물 물동량 이상 감지 분석	44
4.2.1. 이동평균법(MA)을 활용한 국가, 항만, 선사별 물동량 이상 감지 ...	44
4.2.1.1 국가별 물동량 이상 감지 사례	46
4.2.1.2 항만별 물동량 이상 감지 사례	49
4.2.1.3 선사별 물동량 이상 감지 사례	54
4.2.2. ARIMA 분석 방법을 활용한 부산항 전체 물동량 이상 감지	57
4.2.3. 부산항 전체 물동량 이상 감지	60
제5장 결 론	61
5.1. 연구결과의 요약 및 시사점	61
5.2. 연구의 한계점과 과제	63
참고 문헌	65
<국내 문헌>	65
<외국 문헌>	67
<관련 사이트>	67

<표 목차>

[표 2-1] Cisco의 모바일 데이터 트래픽 예측(2013~2018)	12
[표 2-2] 빅데이터의 대표적인 3가지(3V) 구성요소	16
[표 2-3] 시계열 정상성 특징	11
[표 2-4] 빅데이터 활용분야 예시	26
[표 2-5] 물류 사업에서의 빅데이터 활용 사례	27
[표 2-6] 중국 수입환적 물동량 예측 결과	30
[표 2-7] 중국 수출환적 물동량 예측 결과	30
[표 2-8] 부산항 컨테이너 물동량 실제치와 예측치	33
[표 2-9] 부산항 컨테이너 물동량 ARIMA 예측치와 표준오차	33
[표 3-1] 부산항 물동량 처리현황	34
[표 3-2] 년도별 부산항 환적화물 물동량 현황	35
[표 3-3] 부산항 환적 화물 데이터 마트	36
[표 4-1] 2016년 부산항 기준 인도네시아 수입환적 물동량 이상 감지	47
[표 4-2] 2017년 부산항 기준 중국 수입환적 물동량 이상 감지	48
[표 4-3] 2017년 부산항 기준 페루 수입환적 물동량 이상 감지	49
[표 4-4] 2016년 자카르타 자바항 부산항 수입환적 이상 감지	50
[표 4-5] 2016년 자카르타 수라바야항 부산항 수입환적 이상감지	51
[표 4-6] 2016년 중국 칭다오항만 부산항 환적화물 이상 감지	52
[표 4-7] 2016년 중국 닝보항만 부산항 환적화물 이상 감지	53
[표 4-8] 2016년 A선사 부산항 수출환적 이상감지	54
[표 4-9] 2017년 B선사 부산항 수출환적 이상감지	55

〈그림 목차〉

[그림 1-1] 시나리오 분석 및 Process Rolling 과정	9
[그림 2-1] 빅데이터 정의의 범주와 그 효과	10
[그림 2-2] 빅데이터의 출현 배경	11
[그림 2-3] ICT 발전과 빅데이터 출현 배경	12
[그림 2-4] KDD PROCESS	16
[그림 2-5] 데이터마이닝 라이프 사이클	17
[그림 2-6] 자기회귀모형의 자기상관함수, 부분 자기상관 함수	20
[그림 2-7] 이동평균모형의 자기상관함수, 부분 자기상관 함수	21
[그림 2-8] 싱가포르 차세대 무인자동차 항만 TUAS	24
[그림 2-9] 함부르크 항만 스마트 포트	24
[그림 3-1] 부산항 환적 화물 물동량 흐름 패턴 분석	35
[그림 3-2] 부산항 수입 환적 화물 물동량 흐름 패턴 분석	36
[그림 3-3] 부산항 수출 환적 화물 물동량 흐름 패턴 분석	36
[그림 3-4] 부산항 국가별 환적 화물 물동량 흐름 패턴 분석	37
[그림 3-5] 부산항 포트별 환적 화물 물동량 흐름 패턴 분석	37
[그림 3-6] 부산항 터미널별 환적 화물 물동량 흐름 패턴 분석	38
[그림 3-7] 부산항 선박신고별 환적 화물 물동량 흐름 패턴 분석	39
[그림 3-8] 부산항 화물 신고별 환적 화물 물동량 흐름 패턴 분석	39
[그림 4-1] 국가, 포트, 터미널, 선사별 MA분석 프로세스	44
[그림 4-2] 2016년 부산항 기준 인도네시아 수입환적 이상감지	45
[그림 4-3] 2017년 부산항 기준 중국 수입환적 이상감지	46
[그림 4-4] 2016년 부산항 기준 태국 수입환적 이상감지	47
[그림 4-4] 2016년 자카르타 자바항만 이상감지	48
[그림 4-5] 2016년 자카르타 수라바야항만 이상감지	49
[그림 4-6] 2016년 인도네시아 수입 환적 물동량 흐름 패턴 분석	49
[그림 4-7] 2016년 인도네시아 포트별 수입 환적 물동량 흐름 패턴 분석	50

[그림 4-8] 2016년 인도네시아 터미널별 수입 환적 물동량 흐름 패턴 분석 ..	52
[그림 4-9] 2016년 인도네시아 입항 선박별 수입 환적 물동량 흐름 패턴 분석 ..	53
[그림 4-10] 2016년 인도네시아 화물신고별 수입 환적 물동량 흐름 패턴 분석 ..	54
[그림 4-11] 2016년 인도네시아 화물신고별 수입 환적 물동량 이상 감지 ..	55
[그림 4-12] 2016년 네덜란드 수출 환적 포트별 물동량 흐름 패턴 분석	55
[그림 4-13] 2016 네덜란드 수출 환적 터미널별 물동량 흐름 패턴 분석	56
[그림 4-14] 2016 네덜란드 수출 환적 출항 선박신고별 물동량 흐름 패턴 분석 ..	56
[그림 4-15] 2016 네덜란드 수출 환적 출항 선박 사이즈별 흐름 패턴 분석	57
[그림 4-16] 2016 네덜란드 수출 환적 출항 선박 사이즈별 흐름 패턴 분석 ..	59
[그림 4-17] 물동량의 자기상관함수(ACF)와 부분자기상관함수(PACF)	60
[그림 4-18] ARIMA 분석 프로세스	61

<식 목차>

[식 1-1] AR모형	5
[식 1-2] MA모형	8
[식 4-1] MAPE 계산공식	8

제 1 장 서 론

1.1 연구의 배경 및 필요성

부산항의 환적화물은 2011년 7,353천TEU에서 2015년 10,105천TEU로 최근 5년 동안 연평균 8.27%의 증가율을 나타내고 있다. 이는 국내 환적화물 물동량의 95%를 차지하는 양으로 환적화물은 부산항이 거의 모두 처리한다고 봐도 과언이 아니다. 그러나 2016년 9월 한진해운 사태를 비롯해 미국 대통령 트럼프의 보호무역주의, 중국 환율 조작국 지정 등의 경제 기조는 부산항의 환적 활동에 지속적으로 적신호를 보내고 있다. 또한, 세계 교역량의 감소와 선대의 축소 등으로 인한 물량 감소는 환적 화물에 대한 의존도가 상대적으로 높은 부산항에 좋지 않은 영향을 주고 있다.

과거 환적 화물 거점 지정 및 환적 화물 유치 결정권은 선사에 있었으나 앞으로 선사는 영업 손익을 고려해 환적 거점 항만을 결정하고 선사뿐만 아니라 화주의 결정권 또한 높아질 것이다. 그래서 선사나 터미널이 다른 국가, 항만, 터미널에서 부산항으로 환적 되는 물량의 변화가 어떤지 특별히 나타나는 이상치는 없는지 분석하는 자료가 필요하다. 그러나 현재 선사에서 마케팅을 위해 활용하고 있는 자료로는 부산항에 신고된 수입출입 환적 화물의 물량 통계뿐이다. 물론 원시 DATA는 있지만 공개되지 않고 있고, 월별 년별 집계된 데이터가 대부분이다. 이 자료만으로는 의미 있는 국가별, 항만별, 선사별 화물의 흐름, 경향을 분석하기에는 부족하다.

본 논문은 항만산업의 환경 변화에 따라 변화되고 있는 환적화물의 흐름을 빅데이터 기반으로 분석하고, 이를 통해 부산항이 데이터 기반 마케팅 관점에서 환적 화물의 부산항 글로벌 경쟁력 강화 전략을 수립하는데 도움을 주는 것을 목표로 한다.

1.2 연구의 방법

본 연구는 4차 산업혁명의 최신 ICT 기술인 ICBM(IoT, Cloud, Big Data, Mobile) 중 하나인 빅데이터 분석을 통해 데이터에 기반한 부산항 환적화물 유치 방안을 제안할 수 있도록 빅데이터 분석 시나리오를 산출하고 분석 결과를 도출하고자 한다. 이에 관련 공공기관(BPA) 및 유관기관(포워딩, 선사, 운영사, 운송사 등)의 인터뷰를 통해 부산항 환적 화물 흐름 패턴 분석, 부산항 환적화물 이상 감지 등 필요한 요구사항이 반영된 시나리오를 산출하고 빅데이터 분석을 실시하여 부산항 환적화물 유치를 위협하는 경쟁 항만에 실질적인 대응을 할 수 있는 방안을 제안하고자 한다.

분석을 위한 자료는 부산항만공사의 Port-MIS 자료 중 부산항을 경유하는 환적화물 자료 및 를 활용한다. 또한 터미널에서 관리하고 있는 COARRI, CODECO 자료도 일부 포함했다. 시나리오에 대한 분석 과정은 현업 인터뷰 및 아이디어 도출을 통해 분석 요구 사항을 파악 한 후 빅데이터 분석 모델링을 한다. 분석 결과로 비즈니스 인사이트(비용절감, 업무 효율성 증대, 마케팅 향상, 신성장 비즈니스 창출 등)를 도출할 수 있는 가치 및 효과를 파악한다.

시나리오 분석 과정



[그림 1-1] 시나리오 분석 및 Process Rolling 과정

그 다음 분석 대상이 되는 원천데이터를 찾아 수집하고 데이터 마이닝 등 데이터 전처리 과정을 거쳐 데이터를 정제한 후 설계된 빅데이터 분석 모델에 의해 분석한다. 분석된 결과를 현업 인터뷰를 통해 적합성여부를 판단하고 최종 시나리오로 채택한다. 이 때 시나리오의 적합성에 대한 추가적인 피드백 및 Rolling 과정은 계속 진행된다.



제 2 장 이론적 고찰

2.1 빅데이터의 개념 및 특징

2.1.1 빅데이터의 정의

빅데이터의 정의는 다양하다. 빅데이터(Big data)는 방대한 데이터를 일컫는다. 단순히 용량만 방대한 것이 아니라 복잡성과 다양성까지 있어 기존의 데이터 처리 방법이나 툴(Tool)로는 다루기 어려운 데이터셋(Data set)의 집합을 지칭하기도 한다. Mckinsey(2011)의 따르면 빅데이터란 “일반적인 데이터베이스 소프트웨어로는 저장, 관리, 분석할 수 있는 범위를 초과하는 대규모의 데이터이다”라고 정의하고 있다.

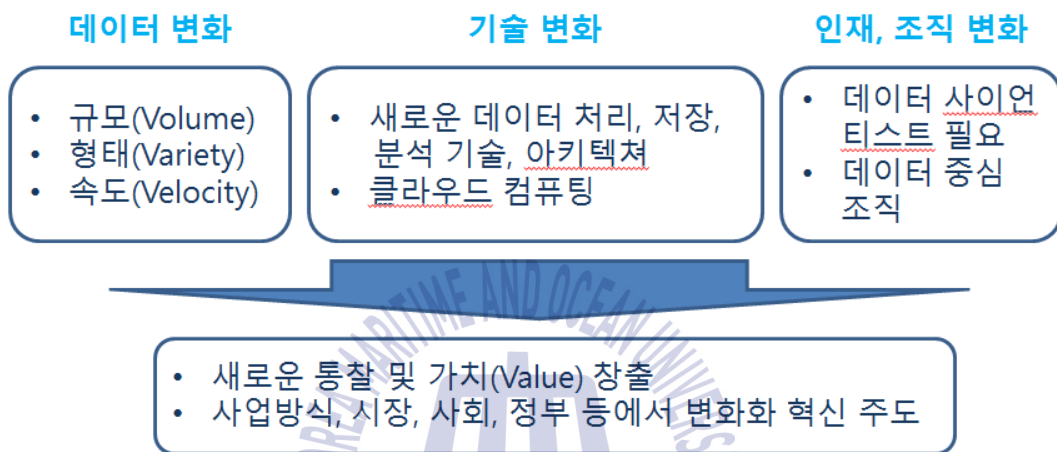
빅데이터 특징으로 흔히들 “3V”를 말하고 있는데, 이는 Gartner Group의 Doug Laney의 정리로 데이터의 양(Volume), 데이터의 유형과 다양성(Variety), 데이터 수집과 처리 속도(Velocity)가 급격히 증가하면서 나타난 현상이다.

빅데이터의 또 다른 정의로 빅데이터 자체 특징보다 이로 인해 나타난 사회·정치·경제·문화적 변화를 포착하기 위한 Mayer-Schonbergerd & Cukier(2013)의 추상적 개념 정의도 있다. 그 내용은 다음과 같다.

“빅데이터란 대용량 데이터를 활용해 새로운 가치나 통찰을 이끌어 내는 일이다. 나아가 이를 활용하여 시장, 기업 및 시민, 정부의 관계 등 많은 분야에 변화를 가져오는 일이다.”

빅데이터 정의를 종합하면 3V로 요약되는 데이터 자체의 특성변화 뿐 아니라 분석 기술의 변화, 나아가 인재, 조직 관점의 변화까지 말할 수 있다.

결국, 빅데이터란 기존의 정형화되고 작은 범위의 데이터 처리 분석으로는 얻을 수 없는 가치와 통찰을 만들어 내는 새로운 방식으로 정의할 수 있겠다.



[그림 2-1] 빅데이터 정의의 범주와 그 효과

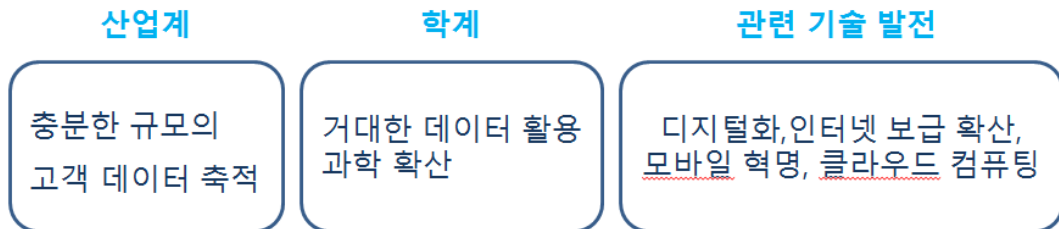
2.1.2 빅데이터의 출현 배경

최근 화두가 되고 있는 4차 산업혁명 기술¹⁾중 하나인 빅데이터 기술은 없었던 현상이 새롭게 등장한 것이 아니라 데이터와 그 기술 차원에서 ‘패러다임의 전환(paradigm shift)’이 일어난 것이다. [그림 2-2]와 같이 산업계와 학계를 둘

1) 인공 지능(AI), 사물 인터넷(IoT), 클라우드 컴퓨팅, 빅데이터, 모바일 등 지능정보기술이 기존 산업과 서비스에 융합되거나 3D 프린팅, 로봇공학, 생명공학, 나노기술 등 여러 분야의 신기술과 결합되어 실세계 모든 제품·서비스를 네트워크로 연결하고 사물을 지능화한 기술.

출처 : [네이버 지식백과] 제4차 산업혁명 [The Fourth Industrial Revolution, 第4次産業革命]

러싼 환경의 변화 관점에서 그 배경을 살펴 볼 필요가 있다.



[그림 2-2] 빅데이터의 출현 배경

기업이 보유한 거대한 양의 데이터는 빅데이터 기술과 접목되면서 온라인, 오프라인을 가리지 않고 TB(Tera byte) 이상의 데이터에서 숨은 가치를 발굴해 새로운 비즈니스로 창출하고 있다. 특히, SNS(Social Network Service)의 등장으로 데이터의 양은 기하급수적으로 증가하고 있으며, 전 세계적으로 생산, 유통되는 디지털 정보량은 2011년 1.8 제타바이트(ZB) 즉 약 2조 기가바이트에 달한다.²⁾ 미국의 네트워크 통신사인 Cisco는 2018년까지 전 세계 모바일 데이터 트래픽이 매달 15.9 엑사 바이트(약 1.6조 기가 바이트)까지 성장할 것이며 2013년 ~ 2018년까지 연평균 61% 정도로 모바일 데이터 트래픽이 증가할 것이라 예측하고 있다.

[표 2-1] Cisco의 모바일 데이터 트래픽 예측(2013~2018)

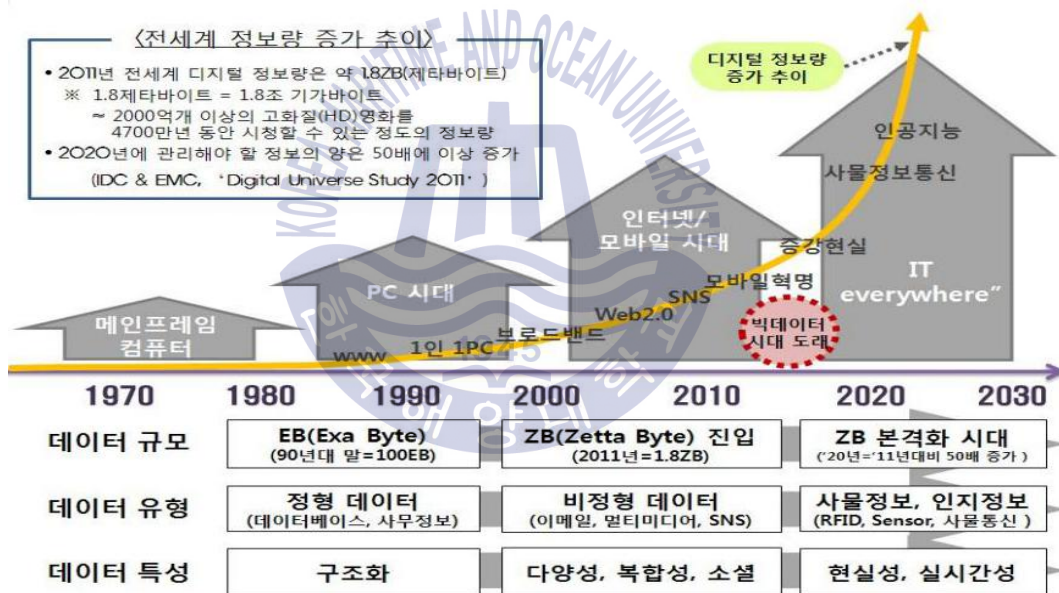
구분	2013	2014	2015	2016	2017	2018
트래픽 양	1.5EB	2.6EB	4.4EB	7.0EB	10.8EB	15.9EB

출처 : Cisco VNI Mobile, 2014

2) 천민경(2014), “ 빅데이터 역량 평가를 위한 참조모델 및 수준진단시스템 개발”, 한양대 석사 학위논문, pp7.

학계에서도 기존 10년이 걸리던 인간 게놈 프로젝트 데이터를 일주일이면 만들 수 있는 기술이 발전했고, NASA의 기후 시뮬레이션 센터에서는 약 32페타바이트의 기후관찰 정보를 활용해 슈퍼컴퓨터에서 시뮬레이션하고 있다.

실제로 일상생활에서 거의 생성되는 데이터들은 일정한 형태가 정해져 있지 않은 비정형 데이터이며 그 개별 데이터 자체로는 아무런 가치가 없는 데이터이다. 그러나 이러한 비정형 데이터도 분석 여하에 따라 소비자의 미래행동을 예측할 수 있고 기업의 위기 대응과 새로운 기회 창출을 동시에 가능하게 할 수 있다.



[그림 2-3] ICT 발전과 빅데이터 출현 배경

*출처: IDC & EMC Digital Universe Study 2011

2.1.3. 빅데이터의 특징

빅데이터의 특징은 일반적으로 3V(Volume, Velocity, Variety)를 기본으로 하고 Value나 Veracity가 추가되어 5V로 나타내는 학자도 있다. IBM은 빅데이터를 3V를 가진 새로운 형태의 데이터로 과거에는 답할 수 없었던 영감을 얻는 기회로 정의하고 있다.³⁾

1) 규모(Volume)

데이터의 발생량을 말하는 것으로 페타바이트, 제타바이트, 요타바이트 이상의 정보를 기준으로 물리적인 크기와 데이터의 속성에도 연관되어 있으며 그것을 처리하는데 어려움이 있는지 아니면 없는지를 의미한다⁴⁾. 오늘날 다양한 문제 해결을 위한 복잡한 패턴을 찾아내기 위해서는 규모가 매우 중요하게 인식되고 있다.

2) 속도(Velocity)

배치 작업을 위한 데이터 처리 속도뿐만 아니라 수많은 사용자의 요청을 실시간으로 처리한 후 처리 결과를 반환해줄 수 있는 기능적인 측면도 포함하고 있다. 데이터의 생성과 처리가 진행되는 속도를 의미 세 가지로 구분된다고 볼 수 있다. 데이터가 발생 후 기업 내의 스토리지에 저장되기까지의 속도와 발생한 데이터의 불필요 부분과 무의미한 부분을 처리하여 가용하게 되는 수준까지

3) 신현신(2016), “빅데이터 기반의 재난정보관리 방안”, 서울시립대석사학위논문 pp6

4) 박두순(2014) “빅데이터 컴퓨팅 기술”, 한빛아카데미, pp19~20

의 속도, 그리고 정제된 데이터를 분석하고 의미를 추출하여 최종 목적을 달성하는 속도까지 말한다. 이러한 속도는 데이터의 접근성과 사용 가능성을 높이는 데 많은 영향을 준다.

3) 다양성(Variety)

빅데이터 분석에 사용되는 데이터는 일반적으로 정형데이터와 비정형데이터로 구분되는데 이중 통일된 구조로 정리하기 어려운 비정형 데이터 분석이 전체 중 90% 이상을 차지한다. 비정형 데이터에는 사진, 동영상 등 기존의 구조화된 데이터가 아닌 다양한 형태의 데이터가 포함된다. 빅데이터를 수집하고 분석할 수 있게 되면서 훨씬 다양한 데이터의 활용이 가능해졌으며, 특히 다양한 구조의 비정형 데이터에서 이전에 생각하지 못했던 인사이트를 얻는 일이 빈번해졌다.

4) 정확성(Veracity)

정확성은 데이터를 얼마나 믿을 수 있는지에 대한 신뢰 수준을 말하는데 보통 데이터가 신뢰에 따라 가치가 부여되기도 한다. 높은 데이터에 대한 품질 유지는 빅데이터의 중요한 요구사항이며 또한 어려운 과제이다. 하지만 최상의 데이터 정제(Data Cleansing)기법을 사용해도 날씨나 경제, 고객의 미래 구매 결정 같은 일부 데이터의 근본적인 불확실성은 쉽게 제거 할 수 없는 것이다. 또한 소셜 네트워크(SNS) 같은 인간 환경에서 생산되는 데이터는 믿기 어렵고, 미래 또한 예측하기 어려워 사람과 자연, 보이지 않는 시장의 힘 등은 빅데이터의 다양한 불확실성을 말해 주고 있다.

5) 가치(Value)

가치는 빅데이터를 저장하려고 IT 인프라 구조 시스템을 구현하는 비용을 말하는데 최종적으로 데이터의 결과가 얼마나 가치를 발휘하는가를 의미한다. 빅데이터의 규모는 엄청나며 대부분은 비정형의 텍스트와 이미지 등으로 구성되어 있다. 이 데이터들은 시간이 지남에 따라 빠르게 전파되면서 변하므로, 그 전체를 파악하고 일정한 패턴을 발견하기가 쉽지 않아 빅데이터 구성요소 중 가치가 무엇보다 중요하다

[표 2-2] 빅데이터의 대표적인 3가지(3V) 구성요소

구분	주요내용
규모의 증가 (Volume)	· 기술적인 발전과 IT의 일상화가 진행되면서 해마다 디지털 정보량이 기하급수적으로 폭증 ⇒ 제타바이트 시대로 진입
속도증가 (Velocity)	· 사물정보(센서, 모니터링), 스트리밍 정보 등 실시간성 정보 증가 · 실시간성으로 인한 데이터 생성, 이동(유통) 속도의 증가 · 대규모 데이터 처리 및 가치 있는 현재정보(실시간) 활용을 위해 데이터 처리 및 분석 속도가 중요
다양성 증가 (Variety)	· 로그기록, 소셜, 위치, 현실데이터 등 데이터의 종류가 증간 · 텍스트 이외의 멀티미디어 등 비정형화된 데이터 유형의 다양화

* 출처 : 정지선(2011), 신가치창출 엔진, 빅데이터의 새로운 가능성과 대응 전략

2.2 빅데이터의 예측 분석 기법

2.2.1 빅데이터 분석 기법

가. KDD(Knowledge Discovery in Databases)

The Knowledge Discovery Process

- **Data Mining v. Knowledge Discovery in Databases (KDD)**

- ▶ DM and KDD are often used interchangeably
- ▶ actually, DM is only part of the KDD process



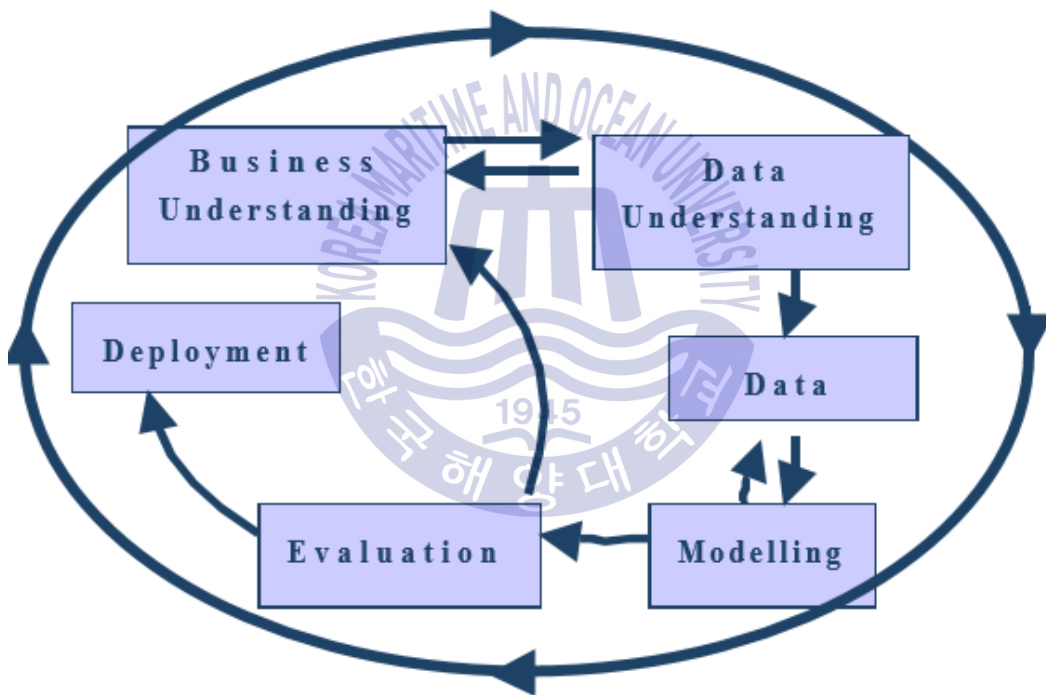
[그림 2-4] KDD PROCESS

*출처: www.cs.uregina.ca/~dbd/831/notes/kdd.html

데이터베이스 내 지식 발견 (KDD)은 데이터 수집에서 유용한 지식을 발견

하는 프로세스이다. 널리 사용되는 데이터 마이닝 기술은 데이터 준비 및 선택, 데이터 정제, 데이터 세트에 대한 사전 지식 통합 및 관찰된 결과의 정확한 해석을 포함하는 프로세스인데 여기서 KDD에서 사용되는 일반적인 방법 중 하나는 일반화, 데이터 값 또는 개념 대체 그 범위에 다른 관련 하위 개념을 포함하는 상위 개념을 사용한다. 비즈니스의 성공적인 존재를 위해 데이터의 기본 패턴을 발견하는 것이 필수적인 것으로 간주된다.⁵⁾

나. CRISP-DM(Cross Industry Standard Process for Data Mining)



[그림 2-5] 데이터마이닝 라이프 사이클

출처: www.ibm.com

라이프사이클 모델은 6개의 단계로 구성되며 단계 사이에는 가장 중요하고

5) 출처 : <https://www.techopedia.com/definition/25827/knowledge-discovery-in-databases-kdd>

빈번한 종속 항목을 표시하는 화살표가 있다. 단계의 순서는 엄격하지 않다. 결국, 대부분의 프로젝트는 필요에 따라 단계 사이를 앞뒤로 이동하며, CRISP-DM 모델은 유연하므로 쉽게 사용자 정의할 수 있다. CRISP-DM을 사용하면 특정 요구사항에 맞는 데이터 마이닝 모델을 작성할 수 있다. 이러한 상황에서 모델링, 평가 및 배포 단계는 데이터 이해 및 준비 단계보다 관련성이 부족할 수 있지만 장기 계획 및 이후의 데이터 마이닝 목적을 위해 이러한 이후 단계 동안 제기된 일부 질문을 고려하는 것은 여전히 중요한 문제이다.⁶⁾

2.2.2 시계열 모형

1) 특징

시간의 흐름에 따라 관측된 자료를 시계열자료(Time-series Data)라 한다. 시계열 분석을 위해서는 먼저, 정상성(Stationary)을 만족해야 한다. 정상성은 일정한 평균, 분산이 시점에 의존하지 않을 것, 공분산은 시차에만 의존하고 시점에는 의존하지 않을 것과 같은 특징을 가지고 있다. 그러나 대부분의 시계열 자료는 정상성을 만족하지 않는다. 따라서 시계열 분석은 정상성을 만족하도록 먼저 정상 시계열 자료를 만드는 작업부터 선행되어야 한다.

정상 시계열 자료를 만들기 위한 첫 번째 작업은 자료로 만들어진 도표를 이용하여 자료의 이상점(Outlier)과 개입(Intervention)을 살펴 정상성을 만족하는지 개략적 추세와 유무를 관찰한다. 이상점의 경우 해당 이상점을 제거하며, 개입의 경우는 회귀 분석을 통해 처리한다. 평균이 일정하지 않은 추세를 보일 경우 차분(Difference)의 과정을 통해 정상 시계열로 바꾸고, 시간에 따른 분산이

6) IBM Knowledge Center

일정하도록 변환(Transformation)의 과정을 통해 정상 시계열로 바꾼다.⁷⁾

차분이란 현 시점의 값에서 전 시점의 값을 빼는 것을 말하는데, 일반적인 차분은 현 시점에서 바로 전 시점 값을 빼는 것, 계절 차분(Seasonal Difference)은 여러 시점 전 자료를 빼는 것을 말한다.

2) 종류

가. AR(자기 회귀 모형, Autoregressive model)

현 시점 자료가 p 시점 전에 유한개의 과거 자료로 설명될 수 있다는 의미이며 AR(p) 모형이라 한다.

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t^{8)}$$

Z_t : 현재 시점의 시계열 자료

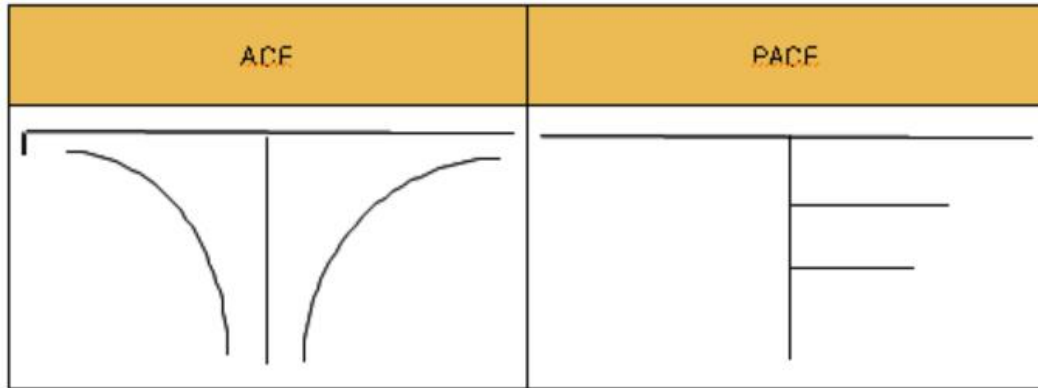
$Z_{t-1}, Z_{t-2}, \dots, Z_{t-p}$: 현재 시점의 시계열 자료

ϕ_p : p 시점이 현재 시점에 얼마나 영향을 주는지 나타내는 모수

a_t : 백색잡음과정(White noise process, 대표적 정상 시계열), 시계열 분석에서 오차항을 의미

7) 한국정보화진흥원 데이터 분석 전문가 가이드 pp 422

8) 한국정보화진흥원 데이터 분석 전문가 가이드 pp.423



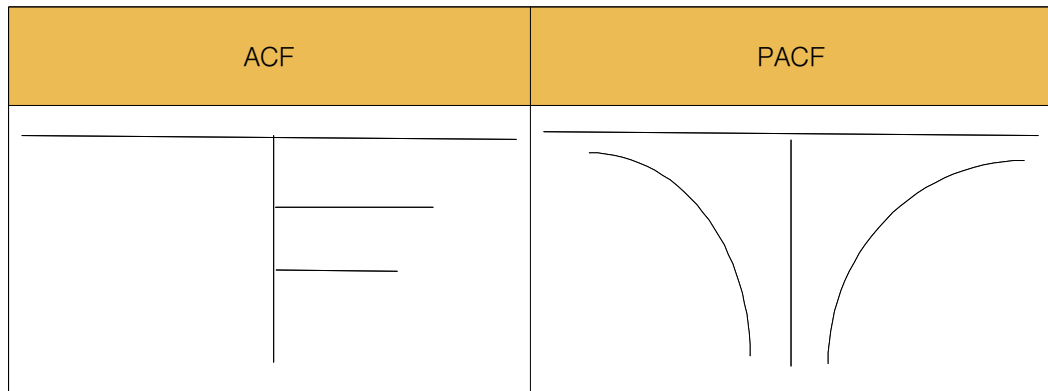
[그림 2-6] 자기회귀모형의 자기상관함수, 부분 자기상관 함수

자기회귀모형인지 판단하기 위한 모형 식별을 위해 자료에서 자기상관 함수 (ACF, Auto-Correlation Function)와 부분자기상관함수(PACF, Partial Auto-Correlation Function)을 이용하여 식별한다. 자기상관함수는 시차가 증가할 때 점차적으로 감소하고, 부분자기상관함수는 p+1 시점 이후 급격히 감소하는 형태로 절단된 형태이며, 이를 AR(p)모형이라고 한다.

나. MA(이동평균 모형, Moving Average Model)

이동 평균 모형은 유한개의 백색잡음의 선형결합이므로 항상 정상성을 만족하게 되어 정상성 가정은 필요 없게 된다.

$$Z_t = a_t - \theta a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_p a_{t-p}$$



[그림 2-7] 이동평균모형의 자기상관함수, 부분 자기상관 함수

이동 평균 모형을 식별하기 위해서는 자기회귀모형과 마찬가지로 자기상관 함수와 부분자기상관함수를 이용하여 식별한다. 이동평균 모형은 자기회귀모형과 반대로 자기상관함수는 $p+1$ 시차 이후 절단된 형태가 되고, 이때를 $MA(p)$ 모형이라고 한다.

다. ARIMA(자기회귀누적이동평균모형)모형

ARIMA 모형은 기본적으로 비정상 시계열 모형이기 때문에 차분이나 변환을 통해 AR모형이나 MA모형, ARMA모형으로 정상화할 수 있다. $ARIMA(p, d, q)$ 모형은 차수 p, d, q 값에 따라 모형의 이름이 달라진다. 차수 p 는 AR모형과 관련이 있고, q 는 MA모형과 관련이 있는 차수이다. d 는 ARIMA에서 ARMA로 정상화할 때 몇 번 차분을 했는지를 의미한다. $d = 0$ 이면 $ARMA(p, q)$ 모형이고 $p = 0$ 이면 $IMA(d, q)$ 모형이라 부른다. $q = 0$ 이면 $ARI(p, d)$ 모형이며, 이를 d 번 차분한 시계열모형이 $AR(p)$ 모형을 따른다.

2.3 해운항만물류 빅데이터 연구 사례

2.3.1 항만 분야 연구 및 활용 사례

가. 싱가포르 TUAS 항만

세계 1위의 환적중심 국가인 싱가포르는 차세대 항만인 TUAS에 2020년부터 2040년까지 모두 65개 선석을 건설하여 모든 시스템을 완전무인자동화시스템으로 운영하는 것으로 확정하였다. 또한 TUAS항은 터미널 물류시스템의 로보틱 기술 이외에도 그린기술, 드론기술, 선박추적 및 정시 입항기술 등 스마트 기술을 동시 실현한 ‘Smart & Green Technology’를 지향하고 있다.

이를 위하여 드론으로 육상에서 해상으로 물품을 전달하고, 선박의 사고 발생 시 피해 부위를 조사한다. 또한, 다목적 부유식 플랫폼을 터미널 인근에 설치하여 선석이 부족할 때 현재의 투묘지 대기 방식이 아닌 출입국 검사, 연료 급유, 선용품 적재등을 사전에 할 수 있도록 한다. 더불어 선박 입출항 편의를 위한 ‘입항 선박 추적시스템 디지털화’, 제출서류 일원화를 위한 ‘Single Sharing Portal’, 입항 선박의 체류 시간을 최소화하고 항만 서비스와 자원투입 최적화를 위한 ‘Just-In-Time Arrival System’ 등을 준비하고 있다. 이러한 혁신적인 물류 시스템은 인공지능을 기반으로 사물 인터넷과 빅데이터 등이 결합하며 가능해진 4차 산업혁명의 산물이다.”⁹⁾

9) 울산항만공사 웹진, 거스를 수 없는 4차산업혁명, 변화의 물결에 올라타라, SK해운 강석환



[그림 2-8] 싱가포르 차세대 무인자동차 항만 TUAS

* 출처: http://www.enewproperty.com/?page_id=685

나. 함부르크 항만

함부르크 항만공사(HPA)역시 SAP, Deutsche Telekom과 함께 함부르크 함부르크 항에 ‘Smart Port Logistics’ 시범 프로젝트 수행, 시스템간의 연계를 이루고 있다. 고질적으로 차량정체 및 물류업체의 항만체류시간 감소를 위해 주요 IT 기업들과 함께 함부르크항에 Smart Port Logistics 시범 프로젝트 시행하고 항만운영사 및 관련 물류기업들의 각기 다른 정보시스템의 정보를 Deutsche Telekom의 Telematic ONE 정보 포털시스템을 통해 통합하고 SAP의 HANA 클라우드 플랫폼을 이용하여 Telematic ONE에 통합된 정보를 클라우드 기반으로 사람이나 차량의 모바일 디바이스에 전달하였다. 해당 프로젝트의 도입하고 빅데이터 분석을 통해 함부르크 항은 기존 항만 인프라를 최적으로 활용, 물류기업의 항만체류시간을 최소화 할 수 있었다.

Smart Port Logistics powered by



[그림 2-9] 함부르크 항만 스마트 포트

* 출처 : Bestfact, 'Smart Logistics at Hamburg Port'

http://www.bestfact.net/wp-content/uploads/2014/02/Bestfact_Quick-info_efreight_3-105_HPA-SmartPortLogistics.pdf

2.3.2 물류 분야 연구 및 활용 사례

나. 물류에서의 빅데이터 연구 및 활용 사례

스마트 물류¹⁰⁾란 운송, 보관, 하역, 시설, 장비 및 물류시스템 등 물류 전 분야에 걸쳐 ICT 기술, 센서, 정보통신 및 제어기술을 활용하여 물류 효율성을 향상시키고 이를 통해 물류비용을 절감하는 것을 목표로 하는 물류로 정의하고 있다.

10) 빅데이터 시대에서의 스마트물류빅데이터 활용 사례 및 동향 고려대 이순교 2016 pp3

지금까지 물류 데이터는 물품의 전달 기본정보, 물품의 무게, 과정마다 기록되는 바코드, QR코드, GPS로부터 물품이 이동되는 경로기록 등 무수한 데이터가 매일매일 쌓여가고 있다. 이러한 데이터들을 최대한 저장해서 활용해 나간다면 물류의 새로운 경쟁력을 높일 수 있다. 바로 이 버려지는 데이터 중에서 의미 있는 정보를 발견하고 기존의 정보와 교차분석을 통해 새로운 인사이트를 발견하는 것이 바로 빅데이터 활용의 핵심이다.

[표 2-4] 빅데이터 활용분야 예시

use case	분야	활용 서비스 예시	설명
Operational Efficiency	Last-mile optimization	Real-time route optimization	배송순서, 교통상황 및 수신자 상태에 따라 배송경로를 동적으로 계산
	Predictive network and capacity planning	Strategic network planning	운송능력 향상을 목적으로 전략적 투자 의사결정을 지원하기 위한 장기 수요예측
		Operational capacity planning	중단기 배송능력 최적화를 위한 인력 및 자원 수요 변동 예측
Customer experience	Customer value management	Customer loyalty management	물류서비스제공자들의 고객 충성도 제고를 위한 고객 데이터 활용
		Continuous service improvement and product innovation	소셜 네트워크 등을 포함한 고객 피드백 등의 활용을 통한 지속적인 서비스 품질 향상 및 새로운 서비스 혁신 추구
	Supply chain risk management	Risk evaluation and resilience planning	공급사슬 상의 공급과잉을 야기하는 다양한 이벤트들의 추적 및 예측

use case	분야	활용 서비스 예시	설명
New Business Models	B2B demand and supply chain forecast	Market intelligence for small and medium-sized enterprises	자체적인 시장분석을 수행하기 어려운 중소기업 대상으로, 예를 들어 선적기록 집계 데이터를 활용한 수요공급 예측 정확도 향상
	Real-time local intelligence	Environmental intelligence	배송 차량에 부착된 다양한 디바이스들(센서, 카메라 등)을 통해 다양한 지역 환경관련 데이터 수집을 하고, 이를 바탕으로 지역 환경 데이터 기반 서비스 제공이 가능

* 출처 : Big Data in Logistics, DHL, 2013

[표 2-5] 물류 사업에서의 빅데이터 활용 사례

구분	데이터	활용 효과 설명
UPS	약 46,000대의 트럭에서 나오는 텔레메틱스 센서들로부터 데이터가 수집	차량의 속도, 방향, 제동, 차량의 성능 등의 정보가 포함. 이 데이터는 배송 기사의 평가 지표로도 활용되지만, UPS 중앙 서버로 수집되어 빅데이터를 구성하여 궁극적으로는 UPS 기사들이 이용하게 되는 주요 경로들을 재설정하는데 쓰임
Schneider National	트럭과 트레일러에 저렴한 비용에 구축 가능한 센서들을 설치해서 데이터를 수집	Schneider가 구현한 연료 정보 센서는 더 나은 연료 최적화를 가능
히타치제작소	업무데이터를 자동적으로 저장하고 이를 분석하는 인공지능(AI) 참고관리 시	일본 히타치제작소는 축적되는 업무데이터를 자동적으로 저장하고 이를 분석하는 인공지능(AI) 참고관리 시스템 개발 하여 이를 통하여 좀 더 효율적인 업무를

	시스템 개발	달성
자라(ZARA)	현재 인기리에 판 매되는 상품의 트 렌드들을 추적하여 패스트푸드처럼 빠 르게 생산하여 공 급하는 전략	미국 매사추세츠 공과대(MIT)와 함께 전 세계 매 장의 판매와 재고에 관한 빅데이터를 실 시간 분석해 수익을 극 대화할 수 있는 '재고 최적 분배 시스 템'

* 출처 : 미래창조과학부, 한국정보화진흥원, 빅데이터센터(2016),“2015년 빅데이터시장
현황조사.”

https://sustainability.ups.com/media/UPS_ORION_2016.pdf.

<http://retailnext.net/en/home/>.

<https://www.youtube.com/watch?v=KkkvOT4G6Ro>.

2.3.3 물동량 변화 예측 연구 사례

가. 부산항에서 중국 환적화물 물동량 분석 및 예측

빅데이터를 이용한 물동량 예측의 사례에 대하여 알아보기로 하자. 부경대학
교 강개선(2013년)¹¹⁾에 따르면 정량적 분석은 물동량 데이터를 바탕으로 2001
년부터 2011년까지 부산항과 환적물동량이 있는 모든 중국 항만의 수입물동
량, 수출물동량, 수입환적물동량 및 수출환적물동량을 기준으로 Pareto분석을
하였다.

또한 정성적 분석은 항만에 관한 중국 및 한국의 각종 신문, 잡지, 학위논문,
학술지 논문, 정부기관의 보고서를 활용하였다. 추가적으로 각 항만공사의 홈
페이지 사이트를 통하여 항만의 입지, 시설, 배후 부지, 해운 및 항만정책, 세
관 특혜의 현황을 기초로 데이터 검증자료에 활용하였다. 정량적 분석과 정성
적 분석을 이용하여 환적물동량을 종속변수로 하여 회귀방정식을 수립을 통한

11) 부산항에서의 중국 환적화물 물동량 분석 및 예측, 부경대, 2013 pp7

물동량 예측방법론을 택하였다.

동 방법론에 따라 물동량 예측한 결과에 따르면 중국의 항만들의 부산에서의 환적물동량에 대한 예측을 위해 각 지역별 항만 배후지역의GDP, 수출입 물동량, 수출입 무역액, 추세변수를 이용하여 회귀분석모델로서 지역별 항만의 환적물동량을 예측하고, 예측된 지역별 물동량의 합으로서 부산항에서의 중국항 환적물동량을 추정하려고 하였다. 여러 변수들을 활용하여 추정한 결과 해당 배후 부지 지역의 수출입 무역액을 이용한 추정 모델이 가장 적합한 것으로 파악되었다. 이는 중국 항만의 수출입화물 경로가 중국항만에서 일본항만으로, 북미항로 및 유럽항로를 통해 운송되는 컨테이너화물의 상당 부분이 부산항에서 환적 되어 이들 지역의 환적화물 물동량은 해당 지역의 대외 수출입 무역액과 직접적으로 연관되기 때문이다.

예측의 방법론으로 환적화물을 수입환적과 수출환적에 대한 추정식으로 나타내었고 수입환적의 실제 부산항과의 물동량에서 각 중국지역(성)에 소재하는 항만들의 수입환적 물동량을 합치고 성별 수입환적 물동량을 천TEU를 단위로 산출하여 앞에 자연대수 형식의 회귀식의 중속변수로 사용한다. 국가 통계국에서 수집하는 화물의 출발지하고 목적지에 의한 통계하는 성별 수출입 무역액 자료를 억 달러를 단위로 정리하여 예측의 방법론으로 환적화물을 회귀식의 설명변수로 대입하였다.¹²⁾ 결과적으로 환적화물의 예측결과는 표 2-6 과 같이 나타났다 상하이를 제외하고는 모든 지역의 수입환적 물동량이 증가하는 추세를 볼수 있다고 표현하였다. 한편 부산항에 수출환적 물동량의 증가 세는 표 2-7 과 같다고 표현하였다.

12) 강개선(2013), “부산항에서의 중국 환적화물 분석 및 예측” 부경대석사학위논문, pp144

[표 2-6] 중국 수입환적 물동량 예측 결과

<3-11> 중국 수입환적 물동량 예측결과

구분	수입환적 물동량(천TEU)					
	실적치		예측치			
	2001	2011	2012	2013	2014	2015
TIANJIN	182	544.9	560.36	611.89	670.13	735.93
SHANDONG	142.93	398.45	389.07	415.06	443.65	475.10
LIAONING	93.38	247.13	270.82	296.02	324.50	356.68
SHANGHAI	146.67	115.89	110.55	107.98	105.46	103.00
GUANGDONG	9.4	70.7	84.55	89.43	94.66	100.29
JIANGSU	5.15	37.98	31.27	33.28	35.45	37.79
기타	101.13	128.49	134.96	146.98	160.20	174.75
합계	680.66	1543.55	1581.57	1700.63	1834.04	1983.54

[표 2-7] 중국 수출환적 물동량 예측 결과

<표 3-13> 중국 수출환적 물동량 예측결과

구분	환적 물동량(천TEU)					
	실적치		예측치			
	2001	2011	2012	2013	2014	2015
TIANJIN	36.38	215.70	210.33	235.28	263.47	295.32
SHANDONG	37.99	209.31	206.53	225.00	245.33	267.68
LIAONING	30.78	104.61	101.84	111.78	123.03	135.73
SHANGHAI	36.14	84.40	83.55	87.66	92.11	96.91
GUANGDONG	6.32	35.86	33.69	36.13	38.75	41.56
JIANGSU	6.14	20.74	19.32	20.83	22.46	24.21
기타	34.67	84.40	87.63	93.20	99.33	106.06
합계	188.41	755.02	742.89	809.89	884.46	967.48

결과적으로 환적화물의 예측결과는 상하이를 제외하고는 모든 지역의 수입환적 물동량이 증가하는 추세를 볼 수 있다고 하였다. 결과적으로 중국의 항만들을 위한 부산항에서의 환적물동량은 계속 증가하는 추세일 것이라고 전망할

수 있다. 다만 상하이의 수입환적 물동량을 -2.33%의 속도로 지속적으로 감소할 것으로 전망된다고 했다. 남중국의 항만들도 북중국 항만의 부산항에서의 환적 물동량보다 작지만 증가세는 유지될 수 있다고 전망된다. 수출환적 물량은 남중국 일부 지역에서 오히려 높은 성장율을 보일 것으로 주목된다. 2015년까지 부산항에서의 중국항 환적 물동량은 수출,수입환적을 합해서 2,950천TEU에 이를 것으로 예측된다고 나타내고 있다.

나. 빅데이터 분석을 통한 물동량 분석 예측 모델에 관한 연구

택배 서비스를 기초로 생산자와 소비자 간의 흐름을 분석하여 택배 물류업에는 차량 정보, 위치정보, 운습도 정보 등 시스템과 데이터의 엄청난 이동을 하고 있으나 이런 비정형 데이터에 대한 수집 이후의 명확한 분석을 할 수 없었다. 빅데이터의 분석을 통해 SCM 상에서 일어나 일련의 데이터의 흐름을 비교 분석하여 물류 고유의 문제를 해결하고 고객의 트렌트 파악과 자원의 효율적 활용에 좋은 매개체가 되고 이것은 향후 고객 관리 및 마케팅 활동과 기업 내부 개선 활동에 영향을 주고 기업의 장기 수익 개선에 기여할 수 있다.

물류시장이 확대되고 있는 추세에 맞추어 질적 성장을 도모하기 위해 택배 물동량의 분석을 통한 연구를 수행하고자 한다. 선행연구를 검토하여 요인 분석을 실시하고, 물동량 분석을 통한 기존 모델의 문제점의 파악을 근간으로 개선모델을 적용하였다. 또한 보완할 수 있는 개선모델을 바탕으로 물동량 분석의 최적화를 이루고자 한다. 이를 통해 물동량 분석 모델 연구에 대한 통합 모형을 제시하고 회귀 분석과 시계열 분석을 반영한 연구모형을 실증적으로 검증하였다.¹³⁾

택배 물동량에 영향을 미치는 조건변수들에 대한 체계를 정하여 42개 변수를

13) 빅데이터 분석을 통한 물동량 분석 예측 모델에 관한 연구 중앙대 이지홍 2017, pp 5~6

발생시켰으며 물동량 분석 모델에 적용하여 물동량의 sub별 영향요인을 도출하였다.

다른 모델의 방법은 회귀분석모형과 시계열 분석모형을 기반으로 수작업 및 경험적 판단에 의존한 배차 예측에서 sub물동량 및 변수가 반영된 데이터 기반의 배차로 개선하여 물동량 분석요인을 파악하여 자원의 효율적 활용하였다. 마지막으로 물동량 분석을 통하여 물동량 분석 모델의 효과를 실증적으로 규명하였다.

본 연구를 통해 실무적 측면에서 물량의 군집화를 수립하고 sub별 특성에 적합한 변수를 생성하였다. 과거 실적에 대한 학습을 통해 물동량 Trend, 추세/계절성을 반영한 예측 모형을 자동으로 적용하였고 최적의 예측 결과를 도출하였다. 2015년 8월 데이터 기준으로 전체 물동량은 월평균 95.2%의 적중률로 신뢰할 수 있는 결과를 취득하였다. 특히, Advanced 모델 물동량은 월평균 93.8%의 적중률, General 모델 물동량은 월평균 95.1%의 적중률, Simple 모델 물동량은 월평균 79.9%의 적중률, Event 모델 물동량은 월평균 71.9%의 적중률로 advanced와 General 모델은 예측 정확도가 KPI 범위에 들었으나 Simple과 Event 모델은 변수가 많아 실 사용에는 한계점이 보였다고 나타내고 있다.

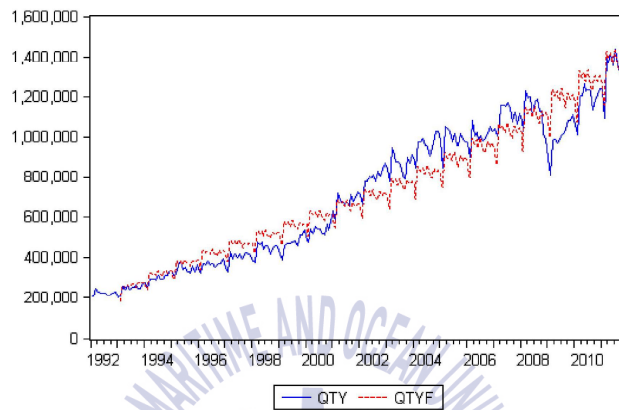
다. 승법계절 ARIMA 모형에 의한 부산항 컨테이너 물동량 추정과 예측

부산항의 물동량 수준의 비정상성과 계절성 혹은 주기성을 고려하여 부산항 물동량의 시계열적 특성을 식별하여 향후 정확한 물동량을 산정하기 위해 승법계절 ARIMA 모형으로 나타내었다. 부산항의 물동량 시계열 특성 즉 비정상성과 주기성 혹은 계절성을 감안하여 부산항 물동량의 시계열적 비정상성과 주기성을 고려하여 연단위(12개월) 단위로 부산항의 물동량 수준을 계절을 승

법한(seasonal multiplicative) ARIMA 모델등 약 7가지를 표시하였다. 그 결과 추정한 그래프는 아래와 같이 나타내었다.

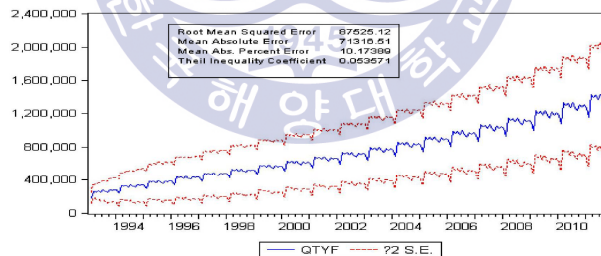
[표 2-8] 부산항 컨테이너 물동량 실제치와 예측치

<그림 5> 부산항의 컨테이너 물동량 실제치와 ARIMA (1,0,1)×(1,0,1)₁₂ 예측치



[표 2-9] 부산항 컨테이너 물동량 ARIMA 예측치와 표준오차

<그림 6> 부산항의 컨테이너 물동량 ARIMA (1,0,1)×(1,0,1)₁₂ 예측치와 표준오차



그 결과로 부산항 연도별 미래 예측 물동량을 산정하였다. 계절 승법한 모형으로 추정했을 때 Akaike information, Schwarz, Hanna-Quin 기준 등으로 보아, 가장 좋은 ARIMA 예측으로 나타났다. 그중 계절 승법한(seasonal multiplicative) ARIMA 모델(1,0,1)×(1,0,1)₁₂로 추정하는 것이 가장 정확도가 높다고 나타내고 있다.

제 3 장 부산항 환적화물 물동량 현황 및 분석

3.1 부산항 환적화물 물동량 현황

부산항 기항 화물 중 PORT-MIS¹⁴⁾에 신고 된 환적 화물에 관한 빅데이터 분석을 이용해 화주 및 선사 마케팅이 가능한 실질적인 데이터 분석을 하고자 한다. 최근 2016년까지의 부산항의 연간 물동량과 연간 환적화물 물동량 등은 <표 3-1>과 같다.

[표 3-1] 부산항 물동량 처리현황

단위 : 천TEU

구분	2011	2012	2013	2014	2015	2016
총계	16,184	17,046	17,686	18,683	19,468	19,432
수출입	8,708	8,808	8,933	9,253	9,363	9,608
환적	7,352	8,147	8,748	9,429	10,105	9,823
연안	124	90	4	-	-	-

* 출처 : 부산항만공사 PORT-MIS

14) 항만 신고와 관련된 Port-MIS는 선박입출항관련업무, 수출입화물 반출입에 관한 업무, 항만시설물에 관한 업무 및 의사결정지원시스템 등을 관리하는 해양수산부 정보시스템임. 포트미스 홈페이지 <http://portmis.go.kr>

[표 3-2] 연도별 부산항 환적화물 물동량 현황

단위: 천TEU

구분	2012	2013	2014	2015	2016
수입환적 물동량	4,126	4,410	4,741	5,064	4,932
수출환적 물동량	4,012	4,325	4,696	5,040	4,893

* 출처 : 부산항만공사 PORT-MIS

3.2 부산항 환적화물 흐름 패턴 분석

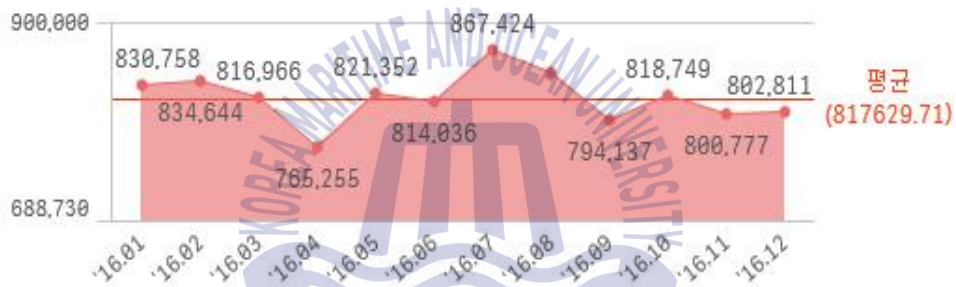
분석에 사용된 자료는 내·외부 정형 데이터를 사용하였다. 내부 자료는 물동량 현황 분석에도 사용된 Port-MIS의 부산항 수입환적 물동량 자료를 사용하였고, 외부 자료로는 선사와 터미널 간에 전송되는 COARRI(Container Discharge/Loading Report Message) EDI 자료를 사용하였다. 부산항을 기항하는 수출입 환적 화물의 패턴을 분석하기 위해 <표 3-3>과 같은 데이터 마트를 구성하였다. 분석에는 최근 2008년 4월 ~ 2017년 4월 10년간의 자료를 사용하였다.

[표 3-3] 부산항 환적 화물 데이터 마트

구분	데이터 마트 항목
수입환적	수입일, 수입요일, 전항지국가, 전항지포트, 전항지터미널, 수입신고업체, 전항지화물, 적컨TEU, 공컨TEU

구분	데이터 마트 항목
수출환적	수출일, 수출요일, 차항지국가, 차항지포트, 차항지터미널, 수출신고업체, 차항지화물, 적컨TEU, 공컨TEU
환적	수입일, 수입요일, 전항지국가, 전항지포트, 전항지터미널, 수입신고업체, 전항지화물, 수출일, 수출요일, 차항지국가, 차항지포트, 차항지터미널, 수출신고업체, 차항지화물, 적컨TEU, 공컨TEU

월별 환적화물 물동량 추이
물동량 (TEU)



[그림 3-1] 부산항 환적 화물 물동량 흐름 패턴 분석

월별 수입환적화물 물동량 추이
물동량 (TEU)



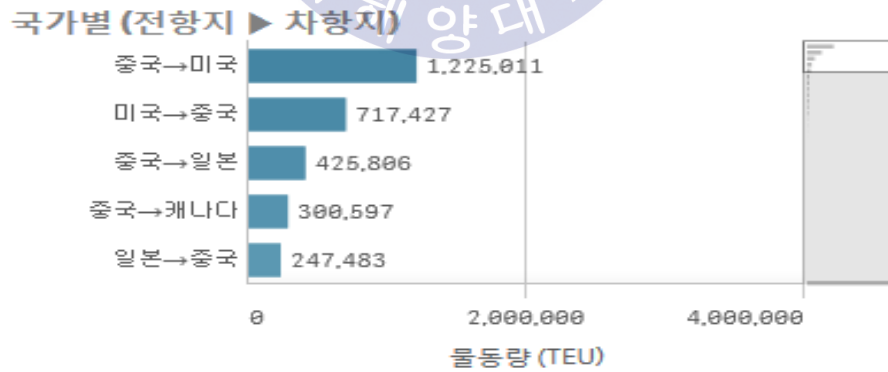
[그림 3-2] 부산항 수입 환적 화물 물동량 흐름 패턴 분석

월별 수출환적화물 물동량 추이
물동량(TEU)



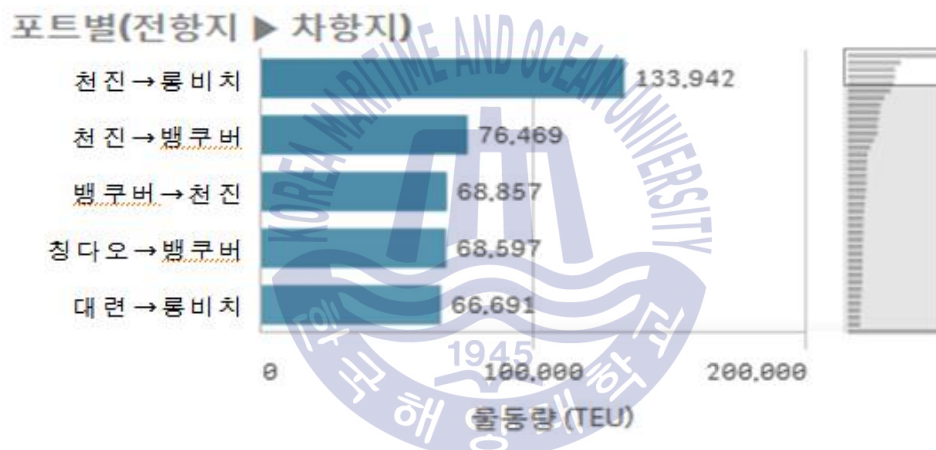
[그림 3-3] 부산항 수출 환적 화물 물동량 흐름 패턴 분석

환적화물의 추이를 분석해 보면 [그림 3-1]과 같다. 그림에서 2016년 4월과 9월 이후 물동량이 감소하는 것을 볼 수 있다. [그림 3-2]와 [그림 3-3]에서도 수출환적 물동량이 2016년 9월 이후 감소하는 것을 발견할 수 있다. 그렇다면 부산항 환적화물의 대부분을 차지하는 그 세부적인 국가는 어디일까?



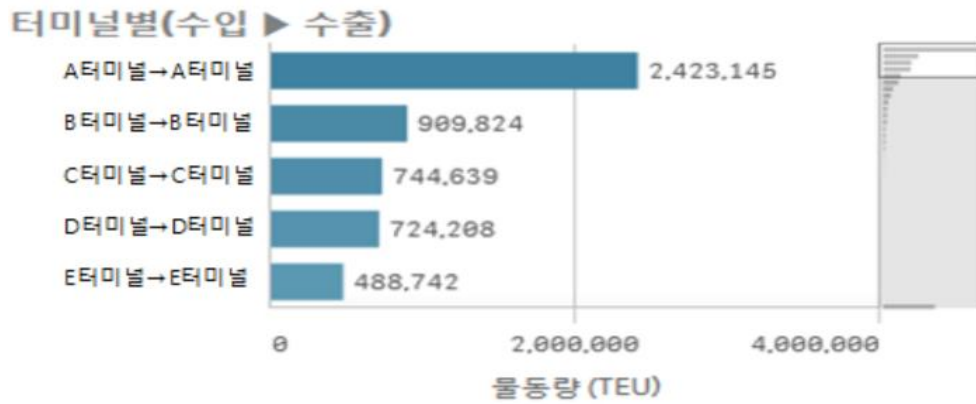
[그림 3-4] 부산항 국가별 환적 화물 물동량 흐름 패턴 분석

부산항을 경유하는 환적화물의 전항지와 차항지의 국가, 항만, 터미널, 입항 신고 업체에 대해서 분석해보면 [그림 3-4]와 같다. 그림에서 보는 바와 같이 국가별 환적화물 물동량은 중국에서 부산항을 거쳐 미국으로 가는 물동량이 가장 많았으며, 미국에서 부산을 거쳐 중국으로 가는 물동량도 두 번째로 많았다. 이는 미국적 글로벌 기업들의 OEM 방식 중국 생산이 많고 이를 다시 미국으로 이동하는 경우가 많다는 점을 시사한다. 이는 부산항 환적화물중 중국 수입환적화물이 대부분을 차지하고 있어 이 부분을 유치하여 유지하는 것이 무엇보다 중요한 점을 시사하고 있다.

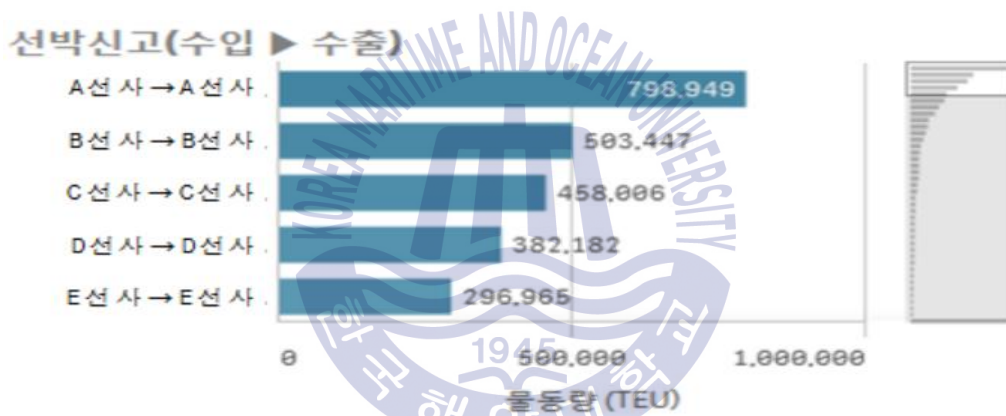


[그림 3-5] 부산항 포트별 환적 화물 물동량 흐름 패턴 분석

항만별로 분석해보면 [그림 3-5]에서 알 수 있듯이 중국 천진항에서 미국 롱비치항으로 이동하는 물동량이 가장 많았다. 그 다음으로는 중국 천진항에서 캐나다 밴쿠버항으로의 물동량이 두 번째, 캐나다 밴쿠버항에서 중국 천진항으로 가 세 번째로 많았다. 마찬가지로 글로벌 기업들이 천진항에서 OEM 방식 제조공장에서 많은 완제품을 미국, 캐나다 등 본사 등으로 이동하는 경우가 많음을 알 수 있다.



[그림 3-6] 부산항 터미널별 환적 화물 물동량 흐름 패턴 분석



[그림 3-7] 부산항 선박신고별 환적 화물 물동량 흐름 패턴 분석

부산항 터미널별 환적 화물은 [그림 3-6]과 같이 A 터미널에서 A 터미널로, B 터미널에서 B 터미널로 C 터미널에서 C 터미널로 이동 순으로 많았다. 이는 부산항 환적은 타부두 TS 보다 자부두 TS가 상대적으로 많음을 알 수 있다.

[그림 3-7]과 같이 선박 신고별 환적화물 물동량 흐름을 보면 수입 환적에서 선박을 신고한 업체가 수출 환적에서도 신고하는 경우가 대부분이다. 화물신고는 주로 수입 환적을 신고한 선박 업체가 수출 선박 신고까지 함께 하는 경우가 많음을 알 수 있다.



[그림 3-8] 부산항 화물 신고별 환적 화물 물동량 흐름 패턴 분석

[그림 3-8]과 같이 화물 신고별 환적화물 물동량 흐름도 수입 환적에서 화물을 신고한 업체가 수출 환적에서도 신고하는 경우가 대부분이다. 화물 신고도 선박업체와 마찬가지로 주로 수입 환적을 신고한 화물 신고 업체가 수출 화물 신고까지 함께 하는 경우가 많음을 알 수 있다.

이러한 패턴 분석을 통해 환적 화물의 흐름의 주요 주체들을 확인할 수 있고, 화물 신고별로 어떤 선사의 물동량이 많고 그 흐름은 어떻게 되는지를 파악할 수 있다. 분석 결과를 보면 부산항의 환적화물 중 중국의 비중은 꾸준히 증가하고 있는 것을 파악할 수 있다. 다만, 이러한 중국 의존도는 최근 글로벌 경기 침체와 중국의 성장 둔화로 인해 부산항의 환적화물의 증가세 둔화나 감소를 초래할 것으로 예상된다.

제 4 장 부산항 환적화물 물동량 이상 감지

4.1 부산항 환적화물 물동량 이상 감지 분석

4.1.1 현행 환적 화물 물동량 이상 감지 방법

선사, 터미널 운영사, 운송사, 포워더 등과 같은 항만 이해관계자와의 수차례 인터뷰를 통해 환적 화물 물동량의 이상 변화 감지를 국가, 항만, 선사, 터미널 별로 미리 파악할 수 있는 분석 자료가 있다면 환적화물 유치 마케팅에 큰 도움이 될 것이라는 의견을 들을 수 있었다.

현재 부산항 환적화물 물동량 예측은 과거 월말 물동량 집계자료와 업무상 경험적 판단을 위주로 산정되고 있다. 환적 화물 이상 감지는 주로 업체별 물동량 통계치에 선사별로 전화나 이메일 등을 통한 경험적 판단에 의한 방법으로 구한 정보를 바탕으로 이루어지고 있다. 물론 인맥을 통한 전화나 이메일로 분석하는 방법은 그동안 축적된 선사 및 터미널 업무 담당자의 경험적 판단에 의한 방법이라 많은 경우 60%이상의 정확성을 가지고 있다고 한다. 그러나 전화나 이메일로 얻는 정보는 실제 분석 결과에 근거하지 않아 객관성이 떨어지며, 수집 과정에서도 많은 시간과 비용을 발생시켜 비효율적이며 단발적이다.

부산항과 관련된 주요 컨테이너 선사는 45개 이상이며, 터미널 8개, 주요 운송사 12개까지 하면 항만관련 주체 수는 60여개를 훨씬 넘어 선다. 따라서 기존과 같은 방법으로는 선사나 터미널, 항만공사 등 해운·항만의 주체들의 이상 변

화에 대해 즉각적인 대응이나 모니터링이 불가능하게 된다.

4.1.2 시계열 모형을 이용한 환적화물 물동량 이상 감지

기존 물동량을 예측하는 방법론에 대한 선행 논문은 양항진(2006), 중력모형을 이용한 부산항의 해상물동량 입출항 패턴 분석, 김정훈(2008), 시계열 모형을 이용한 부산 북항의 물동량 예측, 이재득(2013), 승법계절 ARIMA 모형에 의한 부산항 컨테이너 물동량 추정과 예측 등 부산항 전체에 대한 물동량 예측이 주를 이루었다. 물론 물동량의 중장기적 예측은 단기적인 1년 목표치 설정 및 중·장기적인 항만 건설 계획 수립, 추가적인 항만 건설, 선석 확충에 중요한 결정자료로 쓰이고 있다. 주로 부산항 전체, 부산항 북항 물동량에 대해 관심을 가지고 분석한 결과이다. 그러나 예측에만 초점을 맞추었고 초단기적으로 부산항 전체 및 국가별 항만별 선사별 물동량의 급격한 변화, 이상치에 대해서 분석한 자료는 없었다. 만약 이러한 자료가 있다면 변화 패턴을 분석하여 보다 빠른 사후 대응이 가능할 것이다.

부산항 환적 화물 물동량의 국가별, 항만별, 터미널별, 선사별 이상치를 분석하기 위해서 시계열 모형인 이동평균법(Moving Average, MA)을 이용하였다. 분석 대상은 최근 1년간 월평균 물동량이 1,000TEU를 초과하고 월별 물동량이 있었던 국가, 항만, 터미널, 선사만으로 한정하였다. 부산항 전체 물동량의 이상 감지를 위해서는 ARIMA(Autoregressive Integrated Moving Average) 모형을 사용하였다.

본 연구에서는 이동평균법나 ARIMA 시계열 모형에서 도출된 예측값의 95% 신뢰구간을 통해 관리 상·하한선을 구하고, 월별 물동량이 상·하한 관리선을 벗어나면 이상으로 간주하였다.

4.2 부산항 환적화물 물동량 이상 감지 분석

4.2.1 이동평균법을 활용한 국가, 항만, 선사별 물동량 이상 감지

이동평균법(MA)을 이용한 국가, 항만, 터미널, 선사별 분석 과정은 [그림 4-1]과 같다. 전월 기준 10 년간의 국가/항만/선사/터미널별 수입환적/수출환적/수입/수출 물동량 자료를 이용하여 각각의 최적 이동 기간을 산정하였다. 이동 기간이 정해지면 1차 차분한 자료에 대해 3이상 12이하인 기간을 가지는 모형 중 예측 오차가 가장 작은 모형의 기간을 최종 기간으로 선정하였고, 차분한 자료를 바탕으로 예측 값(과거 차분된 데이터의 평균), 표준편차를 산출하였다.

이동기간을 3에서 12월까지로 한정 한 이유는 기간이 3보다 작으면 1, 2개월의 값만 사용하여 신뢰성이 떨어지고, 12 보다 크면 계절적 요인과 중복되기 때문이다. 산출된 예측 값과 표준편차를 이용하여 관리선(예측 값 $\pm 1.96 \times$ 표준편차)을 산출한 다음 실제 물동량이 관리선을 벗어나면 이상으로 간주하였다.

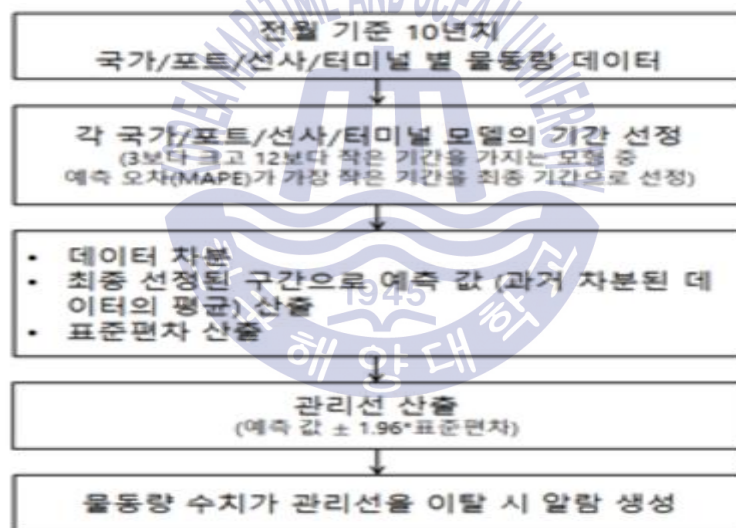
김재경(2015)¹⁵⁾에 따르면 이상점 탐색 시 데이터 확률분포 모형을 가정하고, 각 데이터가 이 확률분포에서 발생할 가능성을 찾았다. 만일 어느 데이터가 가정된 확률분포에서 발생할 가능성이 매우 낮다면 이 데이터는 이상점으로 정의된다. 이 때 가능성은 유의수준 90%, 95%, 99% 등 여러 가지로 정할 수 있다. 통계적 분석 방법 중에 많이 사용하는 방법인 가설 검정에서 제 1종 오류¹⁶⁾를 나타내는 것을 유의 수준 α 이라고 한다. 귀무 가설이 진실인데 이것을 기각시키는 확률이 1%, 5%, 10% 라는 것을 의미하며 검정력은 유의수준 α 와 상호 배타적이라

15) 혼합 극단분포를 이용한 이상점 탐색 연구, 송실대, 김재경(2015)

16) 귀무가설이 진실인데 기각 시킬 확률

할 수 있다. α 가 커지면 검정력이 작아지고, 반대로 α 가 작아지면 검정력이 커지게 된다. 따라서 α 가 너무 작지도 크지도 않는 적절한 경우가 95%이며 대부분의 사회 과학 분야에서 유의수준 95%를 가장 보편적으로 쓰고 있는 이유이다.

나은희(2016)¹⁷⁾에 따르면 이상치를 처리하는 분석 기법 : 귀무가설이 진실인데 기각시킬 확률계열 기법 중 하나인 가중이동평균법으로 유의수준 95% 벗어나면 이상치로 간주하고 있다. 또한 일반적인 사회 과학에서 유의수준 95%를 벗어나면 귀무가설을 기각시키는데 이를 착안하여 95%신뢰구간을 벗어나는 경우를 이상치로 선정하게 되었다.



[그림 4-1] 국가, 포트, 선사별 MA분석 프로세스

예측의 정확도를 평가하는 지표로는 MAPE(Mean Absolute Percent Error)를 사용하였다. MAPE는 실제값(또는 예측값) 대비 예측 오차의 비율의 절대값의

17) 대용량 데이터 분석을 위한 맵리듀스 기반의 이상치 탐지 연구 중앙대,나은희(2016)

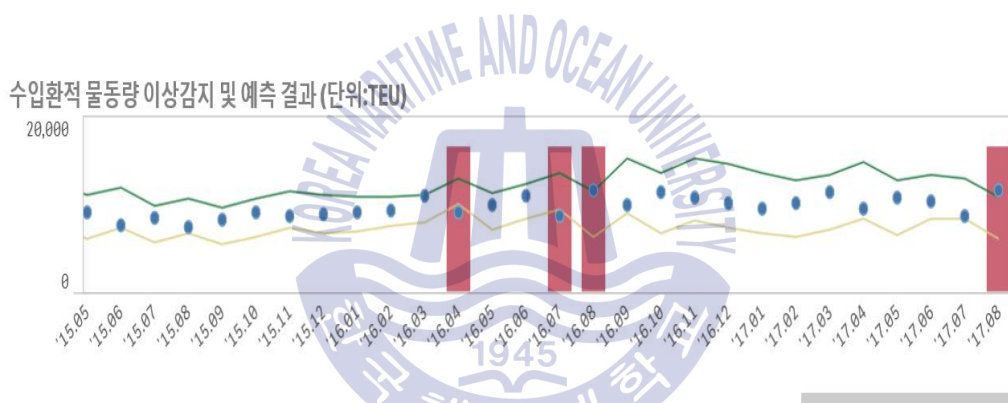
평균이며, 계산공식은 아래와 같다.

[식 4-1] MAPE 계산공식

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i \text{ (or } F_i)} \right|$$

4.2.1.1 국가별 환적 화물 물동량 이상 감지 분석 사례

가. 인도네시아 이상감지



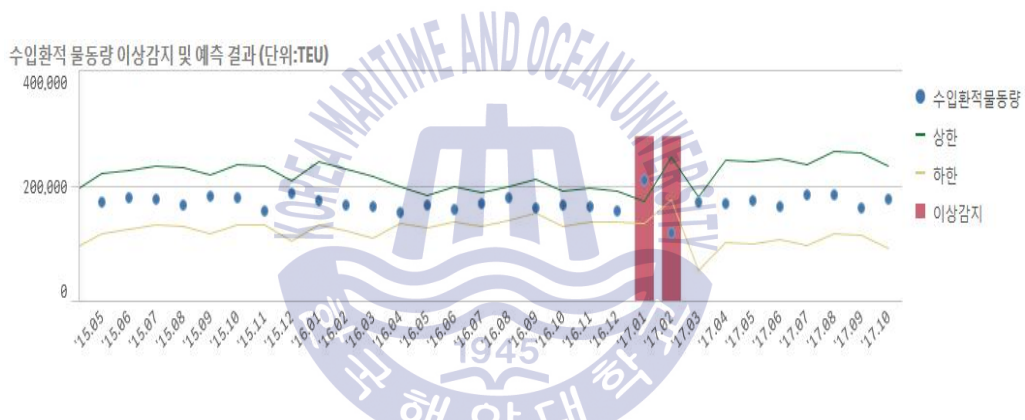
[그림 4-2] 2016년 부산항 기준 인도네시아 수입환적 이상감지

2016년 수입 환적화물의 이상 감지 분석을 통해서 인도네시아의 물동량 이상을 파악할 수 있었다. [그림 4-2]와 같이 2016년 4월, 7월, 8월에 인도네시아 국가의 수입 환적화물 물동량의 이상이 감지되었다. 4월과 7월은 하한 이탈의 부정적인 이상이 감지되었고, 8월은 상한 이탈의 긍정적 이상이 감지되었다. 인도네시아의 주요 수출 품목인 제지 시장에서 여름 휴가를 맞아 소비량이 줄어 생산량도 줄었음을 알 수 있다. 생산량이 줄면 부산으로 환적하는 물량 또한 줄어 들게 된다.

[표 4-1] 2016년 부산항 기준 인도네시아 수입환적 물동량 이상 감지 (단위: TEU)

구분	2016년 4월	2016년 7월	2016년 8월
수입환적	8,803	8,531	11,415
상한	12,575	13,164	11,075
하한	9,640	9,088	5,902

나. 중국 이상감지



[그림 4-3] 2017년 부산항 기준 중국 수입환적 이상감지

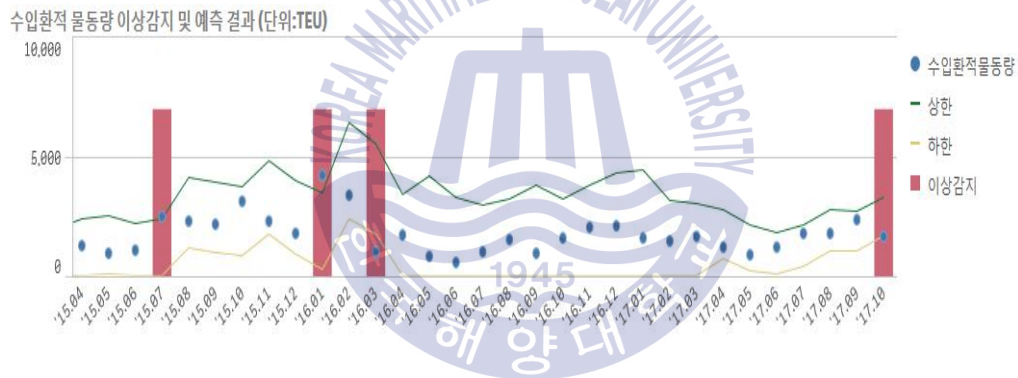
2017년 수입 환적화물의 이상 감지 분석을 통해서 중국의 물동량 이상을 파악할 수 있었다. [그림 4-3]과 같이 2017년 2월 중국 국가의 수입 환적화물 물동량의 이상이 감지되었다. 2017년 1월 상한이탈이 일어나고 2017년 2월 하한 이탈이 연이어 감지되었다. 이는 '사드(THAAD • 고고도미사일방어체계) 보복' 여파로 OEM 방식의 제조로 대미 수출을 하는 경우가 70%로 부산항을 이용한 환적 물량이 대폭 줄었다. 그러나, 2017년 3월에 다시 물동량을 회복하는데 이는 국적 선사가 지속해서 한진해운 물량을 흡수하고 있고, 4월

얼라이언스 재편을 앞두고 주요 선사들이 이용 선박을 교체하려고 부산항에 내린 화물이 늘어난 것으로 분석 된다.

[표 4-2] 2017년 부산항 기준 중국 수입환적 물동량 이상 감지 (단위: TEU)

구분	2017년 1월	2017년 2월	2017년 3월
수입환적	206,543	117,072	169,001
상한	169,841	245,913	177,012
하한	130,946	172,732	49,397

다. 페루 이상감지



[그림 4-4] 2016년 부산항 기준 페루 수입환적 이상감지

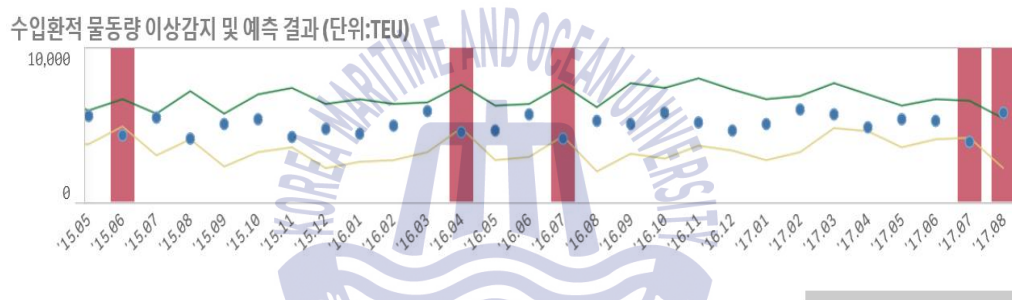
[그림 4-4]와 같이 페루의 2016년 수입환적 물동량 이상 감지 결과를 보면 1월에는 긍정적인 상한선 이탈을 보이다가 3월이 되면 하한선 이탈을 보이고 있다. 페루의 주요 대중국 수출품인 비철 금속류의 중국 수요 감소가 비철 금속 하락세로 이어지고 비철 금속 수출 물량을 감소시키는 결과를 만든 것이다.

[표 4-3] 2017년 부산항 기준 페루 수입환적 물동량 이상 감지 (단위: TEU)

구분	2016년 1월	2016년 2월	2016년 3월
수입환적	4,190	3,327	989
상한	3,416	6,329	5,415
하한	227	2,345	1,682

4.2.1.2 항만별 환적 화물 물동량 이상 감지 분석 사례

가. 자바, 수라바야항만 이상감지



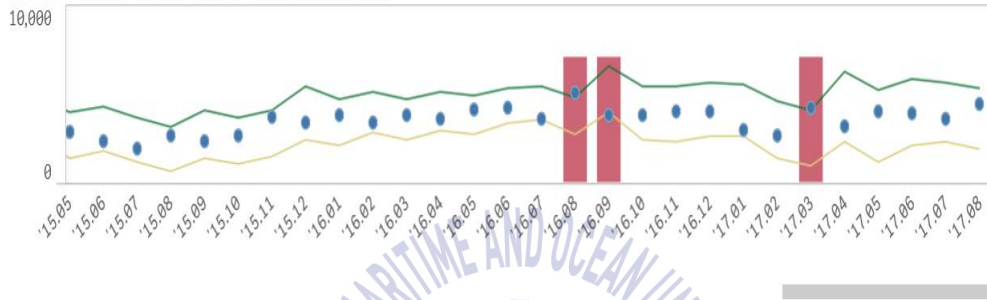
[그림 4-5] 2016년 자카르타 자바항 부산항 수입환적 이상감지

항만별 이상 감지를 살펴보면 인도네시아 항만 중 자카르타 자바항, 수라바야항에서 이상이 감지되었다. [그림 4-5]와 같이 자바항에서 4월과 7월에 상한을 넘어선 물동량 이상이 감지되었다. 여름 휴가를 맞아 인도네시아에서 물동량 하한의 이상 감지를 보였던 것과 마찬가지로 자카르타 자바항, 수라바야항에서도 여름 비수기에는 제지 수요가 낮고 수요가 낮다 보니 수출량에까지 영향을 미쳤다.

[표 4-4] 2016년 자카르타 자바항 부산항 수입환적 이상 감지(단위: TEU)

구분	2016년 4월	2016년 7월
수입환적	4,398	3,980
상한	7,483	7,429
하한	4,648	4,704

수입환적 물동량 이상감지 및 예측 결과 (단위:TEU)



[그림 4-6] 2016년 자카르타 수라바야항 부산항 수입환적 이상감지

[그림 4-6]에서 보듯 수라바야항에서는 2015년 8월에 환적 물동량이 상한 이탈하는 긍정적 이상치가 감지되었고, 9월에는 환적 물동량이 하한을 하회하는 부정적 이상치가 감지되었다.

[표 4-5] 2016년 자카르타 수라바야항 부산항 수입환적 이상감지(단위: TEU)

구분	2016년 8월	2016년 9월
수입환적	4,957	3,700
상한	4,663	6,376
하한	2,510	3,838

추가적으로 이 시기 인도네시아 국가의 수입 환적 물동량 흐름 패턴을 살펴보

면 [그림 4-7]와 같다. 4월과 7월에 평균에 못 미치는 하락세를 확인할 수 있다.

월별 수입환적화물 물동량 추이

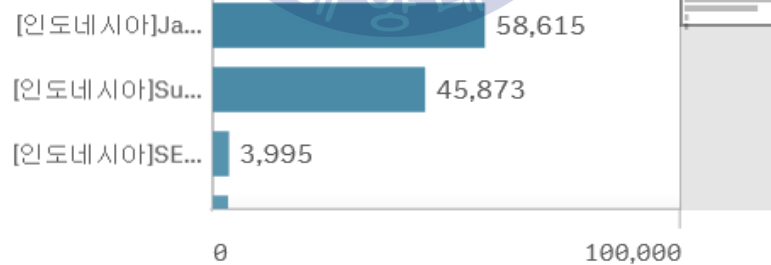
물동량(TEU)



[그림 4-7] 2016년 인도네시아 부산항 수입 환적 물동량 흐름 패턴 분석

[그림 4-6]과 [그림 4-7]과 같이 인도네시아 국가에서 부산항으로 환적 되는 전항지 항만은 자카르타 자바와 자카르타 수라바야이고 해당 수입 터미널 순위도 알 수 있다. 또한 [그림 4-8], [그림 4-9]와 같이 입항신고 선사와 화물신고 선사도 알 수 있다.

전항지

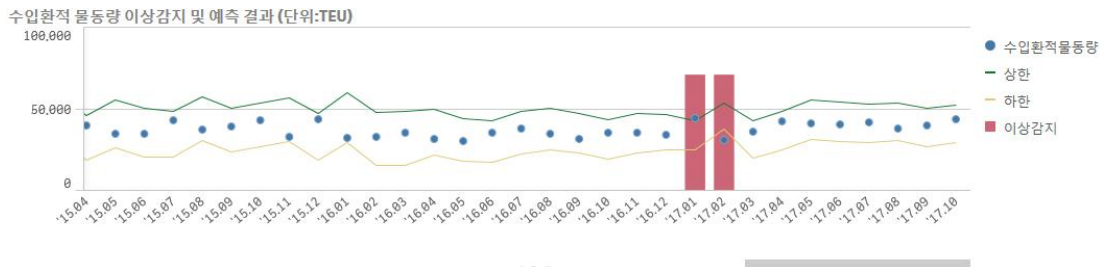


[그림 4-6] 2016년 인도네시아 부산항 항만별 수입환적 물동량

수입 환적과 마찬가지로 수출 환적 물동량도 분석이 가능하다. 2016년 인도네

시아 수출 환적 이상 감지 결과를 보면 [그림 4-10]과 같은데, 3월, 5월, 9월에 각각 이상치가 감지되었다. 해당 이상감지 값을 살펴보면 <표 4-4>와 같이 3월과 9월에 하한 이탈이 있었고, 5월에 상한 이탈로 긍정적인 이상감지를 보였다.

나. 칭다오, Ningbo항만 이상감지

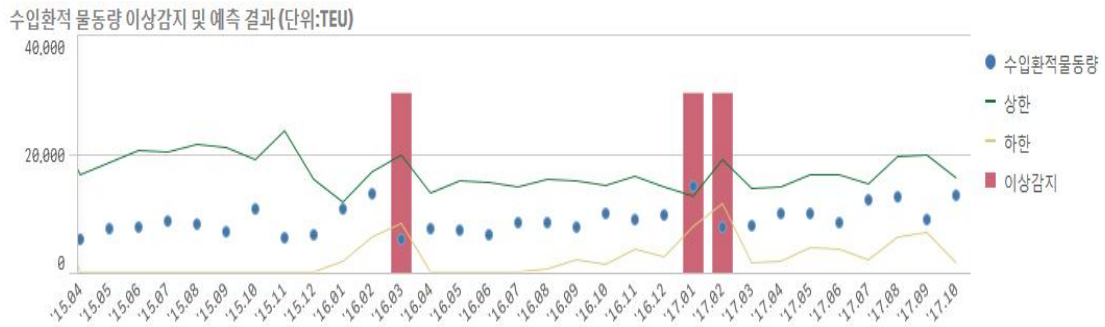


[그림 4-8] 2016년 중국 칭다오항 부산항 수입환적 이상감지

2017년 수입 환적화물의 이상 감지 분석을 통해서 중국의 물동량 이상을 파악할 수 있었다. [그림 4-8]과 같이 2017년 2월 중국 항만 중 상하이항만의 수입 환적화물 물동량의 이상이 감지되었다. 이는 2017년 2월 중국 화물 물량중 대다수를 차지하는 칭다오항에서 부산항 수입환적 화물 또한 사드 보복 조치로 급감했음을 알 수 있다.

[표 4-6] 2016년 중국 칭다오항만 부산항 수입환적 이상 감지(단위: TEU)

구분	2017년 1월	2017년 2월
수입환적	43,823	30,449
상한	41,415	51,826
하한	24,102	36,826



[그림 4-9] 2016년 중국 Ningbo항 부산항 수입환적 이상감지

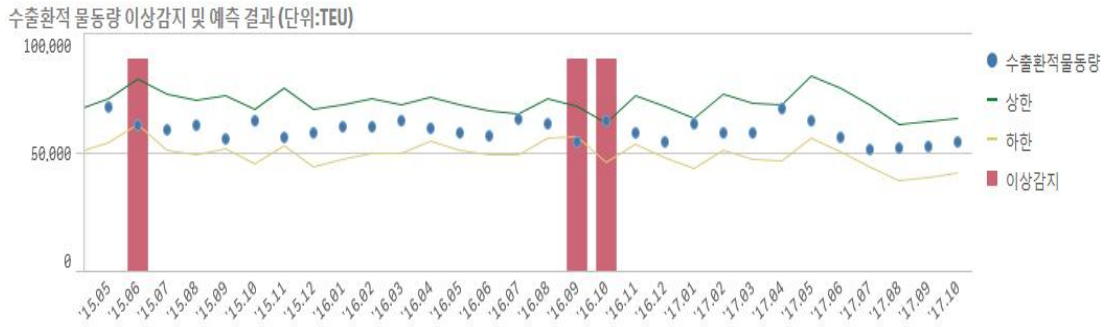
[표 4-7] 2016년 중국 Ningbo항만 부산항 수입환적 이상 감지(단위: TEU)

구분	2017년 1월	2017년 2월
수입환적	14,320	7,545
상한	12,677	18,761
하한	7,623	11,439

상하이항과 마찬가지로 중국 Ningbo항에 대해서도 [그림 4-9]와 같이 2017년 1월과 2월 각각 Ningbo항에서 부산항으로 수입되는 환적화물 물동량의 이상이 감지되었다. 2017년 1월에는 상한을 넘어서는 상한 이상감지가, 2017년 2월에는 하한 아래로 떨어지는 하한 이상 감지가 일어났다. 이는 2017년 2월 중국 화물 물량중 미국 글로벌 기업으로 수출하는 화물이 많이 차지하는 Ningbo항에서 부산항 수입환적 화물 또한 사드 보복 조치로 급감했음을 알 수 있다.

4.2.1.3 선사별 환적 화물 물동량 이상 감지 분석 사례

가. A선사 이상감지



[그림 4-10] 2016년 A선사 부산항 수출환적 이상감지

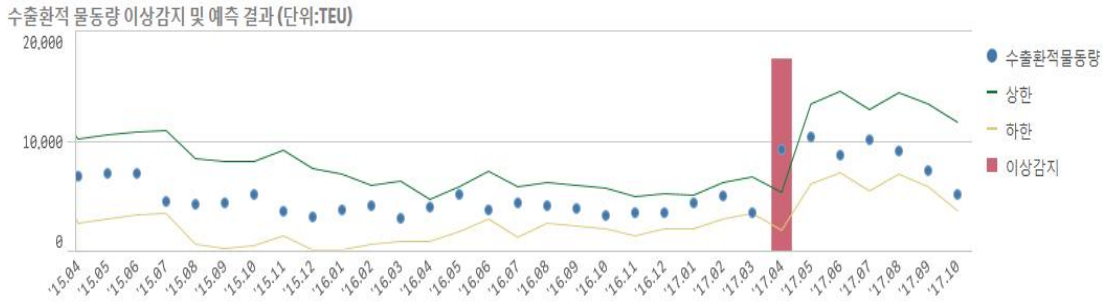
[표 4-8] 2016년 A선사 부산항 수출환적 이상감지(단위: TEU)

구분	2016년 9월	2016년 10월
수입환적	53,857	62,471
상한	68,509	61,695
하한	55,666	45,178

이번에는 선사별 이상감지를 살펴 본다. 세계 5위안에 드는 선사인 A선사의 2017년 9월 부산항 수출 환적 화물 이상감지를 살펴본 결과 2017년 8월부터 약간의 감소 추세를 보이다가 9월에 53,857TEU로 하한 이상감지가 되었다. 이는 2016년 6월말 발생하였던 랜섬 웨어 영향으로 2017년 7월초 일시적인 물류 대란이 일어났고, 그 여파로 9월 수출환적 물량 감소에 영향을 준 것으로 분석된다. 랜섬 웨어 영향으로 그 자회사인 APM 터미널의 일부 마비 및 제한적 운행은 화주들로 하여금 A선사가 아닌 다른 선사를 이용하게 만든 것이다.

나. B선사 이상감지

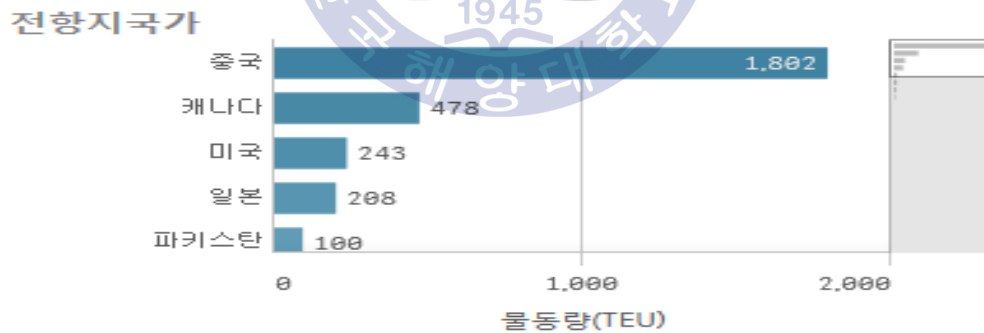
다음으로 북미 와 동북아시아 환적화물 물량이 많은 B 선사의 이상감지를 살펴 보았다.



[그림 4-11] 2016년 B선사 부산항 수출환적 이상감지

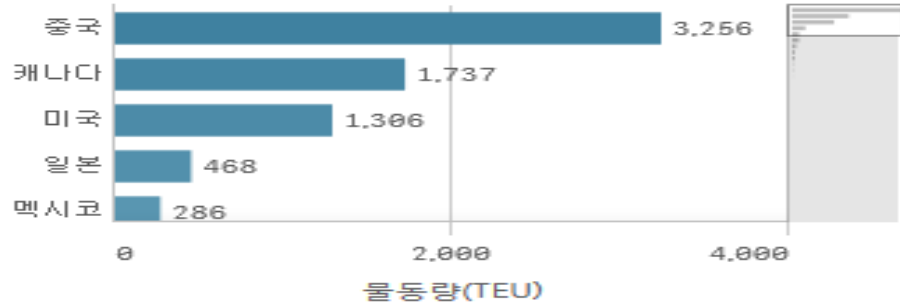
[표 4-9] 2017년 B선사 부산항 수출환적 이상감지(단위: TEU)

구분	2017년 3월	2017년 4월
수입환적	3,433	9,154
상한	6,576	5,221
하한	3,265	1,738



[그림 4-12] 2016년 B선사 2017년 3월 부산항 수입환적 전항지 국가

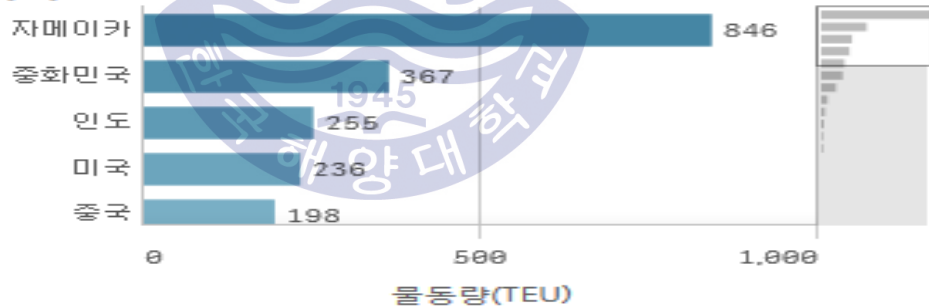
전항지국가



[그림 4-13] 2016년 B선사 2017년 4월 부산항 수입환적 전항지 국가

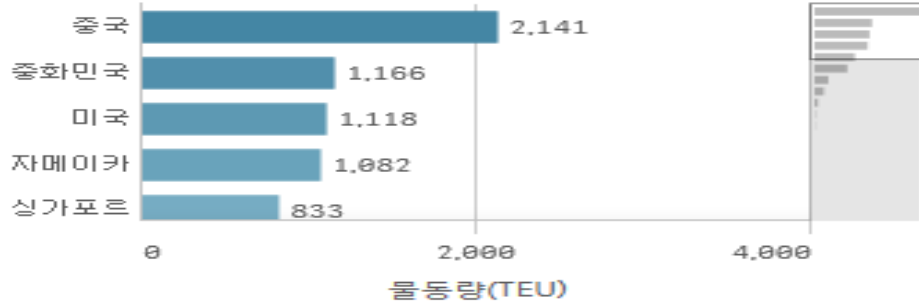
그림 [4-11]과 [4-12]와 같이 B선사의 중국에서 부산항 수입환적 물량이 1,082TEU에서 3,256TEU로 3배나 증가한 것을 알 수 있다. B 선사의 캐나다에서 부산항 수입환적 물량도 478TEU에서 1,737TEU로 3.6배 증가한 것을 알 수 있다.

차항지국가



[그림 4-14] 2016년 B선사 2017년 3월 부산항 수출환적 차항지 국가

차항지국가



[그림 4-15] 2016년 B선사 2017년 4월 부산항 수출환적 차항지 국가

그림 [4-13]과 [4-14]와 같이 B선사의 중국물량이 565TEU에서 3,307TEU로 5.8배나 증가한 것을 알 수 있다. 또한 주요 국가였던 자메이카 물량도 846TEU에서 1,082TEU로 1.3배 증가하였다. 이는 캐나다 주요 아시아 수출 품목인 펄프의 운송이 봄철을 맞아 증가한 것으로 분석된다.

4.2.2. 부산항 전체 환적 화물 물동량 이상 감지 분석

앞에서 분석한 국가, 항만, 터미널, 선사별 이상 감지는 이동평균법(MA) 모형을 사용하였지만 부산항 전체 물동량에 대한 이상 감지는 그보다 더 모수가 많은 ARIMA 모형을 사용하여 분석을 시행하였다.

ARIAM 모형도 전월 기준으로 과거 10년간의 월별 부산항 수입 환적 화물, 수출 환적 화물, 수입화물, 수출화물 물동량 이상 발생 정보를 도출한 다음 이동평균법과 마찬가지로 식별(Identification), 추정(Estimation), 진단(Diagnostic), 예측(Forecasting)의 4단계를 거쳐 분석할 모형을 선정하였다. 식별 단계에서는 주어진 시계열 자료를 바탕으로 시계열이 계절적 특성을 지니고 있는지 여부를 판단

한다. 이와 함께 ARIMA(p,d,q)(P,D,Q)s 모형의 'p', 'd', 'q', 'P', 'D', 'Q', 's'의 가능한 값들을 추정하고, 추정 단계에서는 앞선 식별단계를 통해 알아낸 'p', 'd', 'q', 'P', 'D', 'Q', 's' 값들 가운데, 통계적으로 가장 적절한 값을 결정하고 이와 함께 각 항들에 대한 모수를 추정하였다.

각 항들의 모수를 추정하는 대표적인 방법으로는 최대우도 추정법¹⁸⁾(Maximum Likelihood Estimation), 최소제곱 추정법¹⁹⁾(Least Squares Estimation), 적률 추정법(Method of Moment Estimation)²⁰⁾ 등이 있는데, 본 논문에서는 최대우도 추정법을 사용하여 모수를 추정하였다.

진단 단계에서는 앞서 추정 단계에서 추정된 모형이 통계적으로 유의한지 여부와 모형이 예측 모형으로써 적절한지 여부를 판단하였다. 모형 진단은 주로 잔차분석(Residual Analysis), 과적합진단(Overfitting Diagnostics)²¹⁾ 등을 사용하는데, 본 연구에서는 잔차 분석을 선택하여 모형을 진단하였다. 잔차 분석은 잔차의 자기상관함수와 Portmanteau 검정²²⁾을 시행하는데, Portmanteau 검정에 사용

18) 최대우도 추정법 : 영국의 유전학자이자 통계학자인 피셔(Ronald Aylmer Fisher)에 의해 고안된 기법으로, "관측 결과를 가장 잘 설명해줄 수 있는 추정량을 모수의 추정량을 택하는 방법" .(송문섭, 허문열, 『수리통계학』, 박영사(2002) pp165)

19) 최소제곱 추정법 : 최소제곱 추정법은 "측정값을 기초로 해서 적당한 제곱합을 만들고 그것을 최소로 하는 값을 구하여 측정결과를 처리하는 방법" .(송문섭, 허문열, 『수리통계학』, 박영사(2002) pp160)

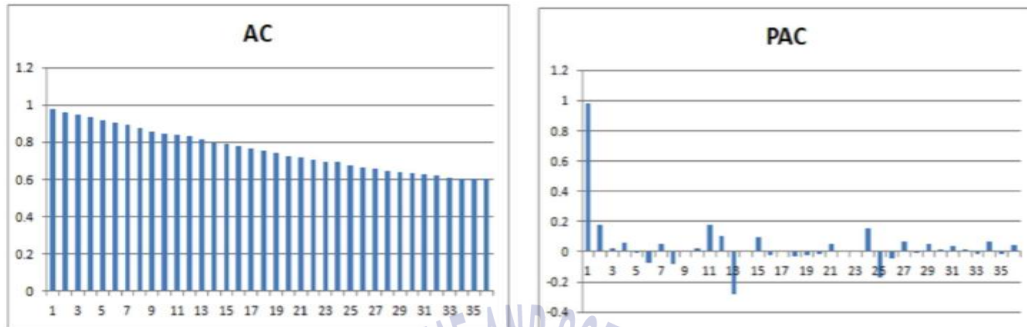
20) 적률 추정법 : "적률(moment)은 '임의의 점에 대한 가중 효과 내지 가중치'를 의미한다고 하는데, 확률분포에 있어서는 평균과 분포, 왜도와 첨도 등을 일반화시킨 것" .(송문섭, 허문열, 『수리통계학』, 박영사(2002) pp162)

21) 과적합진단(Overfitting Diagnostics):모델이 과거의 데이터를 너무 과하게 설명한 나머지 실제 변수들 간의 관계를 잘못 설명하게 되는 경우 모델이 실제 변수들 간의 관계보다는 과거 학습데이터 (training data)의 노이즈를 설명하게 되는 경우인지를 진단(송문섭, 허문열, 『수리통계학』, 박영사(2002) pp151)

22) Portmanteau 검정 :일반적인 시계열의 자기상관성을 검증하는 Ljung-Box test, Box-Pierce test를 보완해 ARIMA모형의 잔차의 자기상관성을 검증하는 방법

[출처] [Box-Jenkins ARIMA모형에 의한 KOSPI와 Historical Volatility 이슈](#)

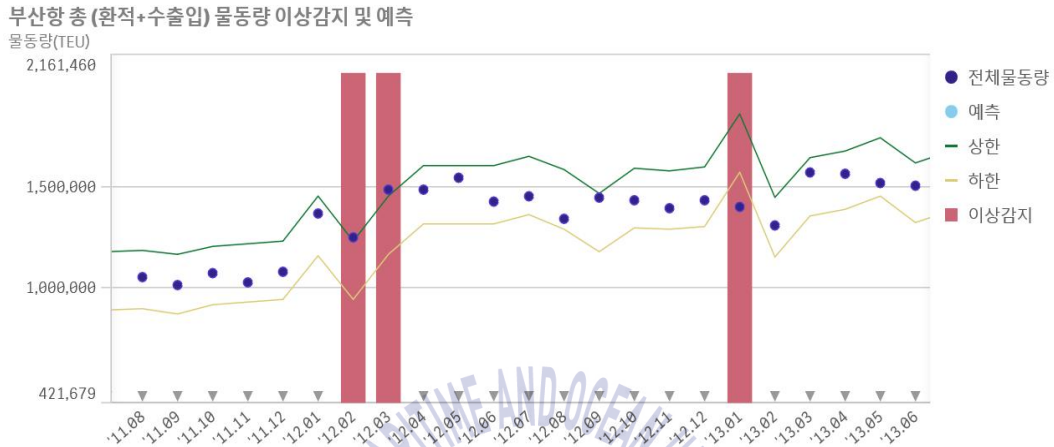
되는 Ljung-Box 통계량은 일정 기간 동안 일련의 관측치가 랜덤이고 독립적인지 여부를 알려준다. 만약 관측치가 독립적이지 않으면 한 관측치가 k 단위 뒤의 다른 관측치와 상관될 수 있다. 이러한 관계를 자기 상관이라고 한다.



[그림 4-16] 물동량의 자기상관함수(ACF)와 부분자기상관함수(PACF)

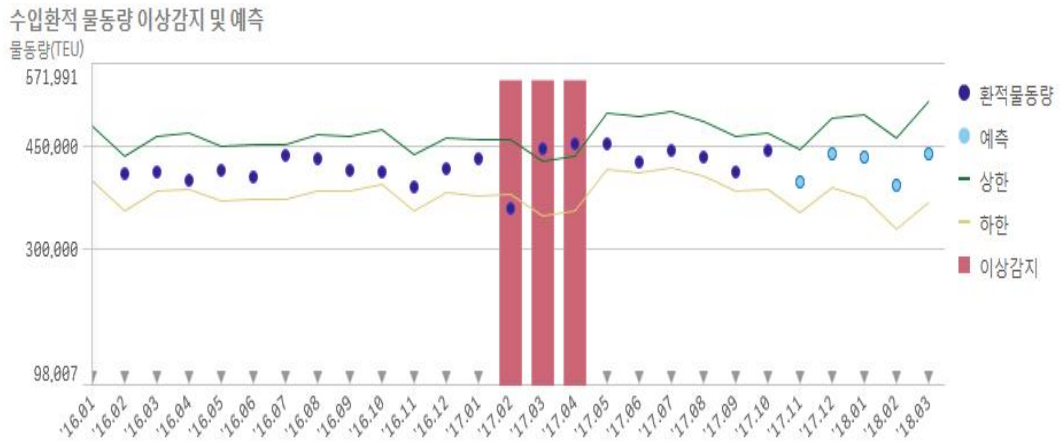
부산항 물동량의 자기상관함수(AC)와 부분자기상관함수(PAC)는 [그림 4-16]과 같다. 둘 다 정상성을 만족하고 있다. 자료의 평균이 일정하지 않은 추세를 보여 차분을 통해 정상 시계열로 바꾸는 과정을 거쳤다. 그리고 Portmanteau 검정을 실시하여 잔차의 독립성을 검증하였고 모형이 적합하다고 판단하였다. 도출된 모형을 이용해 장래수요를 예측하였고, 상하한 관리선을 구하기 위해 예측 값의 95% 신뢰구간을 사용하였다.

4.2.3. 부산항 전체 물동량 이상 감지



[그림 4-17] 2012년 부산항 전체 월별 부산항 물동량 이상감지

2012년 부산항 전체 월별 부산항 물동량 이상감지값을 살펴보면 [그림 4-17]와 같이 2012년 2월과 3월엔 상한 이탈이 있었고, 2013년 1월엔 하한이탈로 부정적인 이상감지를 보였다. 그 원인에 대해서는 당시 해당 국가의 상황, 인접 경쟁항의 상황 등 여러 가지 변수가 있을 수 있어, 수입, 수출, 수입환적, 수출환적을 분리하고, 앞서 분석한 국가, 항만, 선사별 이상 감지를 이용해 세부 상황을 확인하면 된다.



[그림 4-18] 2017년 부산항 전체 월별 부산항 물동량 이상감지

2017년 2월은 중국의 사드 보복의 영향으로 중국발 미국 환적물량이 대거 감소하였으며, 이는 부산항 환적화물 전체에도 영향을 미쳤다. 부산항의 환적화물의 대중국 의존도가 높고 이는 상대적으로 급부상하고 있는 베트남, 태국, 인도네시아 등의 부산항 수입 환적 늘일 수 있는 방안을 세우는 것이 시급하다 하겠다.

제 5 장 결 론

5.1 연구결과의 요약 및 시사점

부산항의 환적화물은 2015년 10,105천 TEU로 최근 5년 동안 연평균 8.27%의 증가율을 보이고 있다. 그러나, 2016년 9월 한진해운 사태, 미국 대통령 트럼프의 보호무역주의, 중국 환율 조작국 지정 등 부산항 환적 화물 물동량의 감소세를 지속시키고 있는 환적 화물 거점 지정 및 환적 화물 유치 결정권은 대부분 선사에게 결정권이 있다. 그러나, 선사에서 다음해 목표치를 정할 때 경험적인 과거 실적을 바탕으로 그 보다 몇% 정도 올려서 정하고 있다.

부산항 환적 화물 물동량을 유치하기 위한 기존 연구에서는 주로 환적 화물 물동량 예측을 위주로 하였다. 그러나 본 연구에서는 빅데이터 분석을 통한 체계적인 연구를 통하여 빅데이터 시대에 데이터에 기반을 둔 물동량 이상치 감지 자료를 만들어 기존 수기 및 현장의 경험적인 판단 방법만으로 운영되는 한계점을 개선하고 물동량 이상 변화에 대한 신뢰성과 빠른 피드백을 줄 수 있을 것이다.

분석방법으로 선택된 것은 먼저 이동평균법(MA)을 활용한 국가, 항만, 선사별 물동량 이상 감지와 ARIMA 분석 방법을 활용한 부산항 전체 물동량 이상 감지이다. 이동평균법나 ARIMA 시계열 모형에서 도출된 예측값의 95% 신뢰구간을 통해 관리 상·하한선을 구하고, 월별 물동량이 상·하한 관리선을 벗어나면 이상으로 간주하였다. 일반적인 사회 과학에서 유의수준 95%를 벗어나면 귀무가설을 기각시키는데 이를 착안하여 95%신뢰구간을 벗어나는 경우를 이상치로 선정하게 되었다. 물동량 이상 발생 정보를 도출한 다음 식별(Identification), 추정(Estimation), 진단(Diagnostic), 예측(Forecasting)의 4단계를 거쳐 분석할 모형을

선정하고 분석을 시행하였다. 그 결과 국가, 항만, 선사별로 이상치를 감지할 수 있었고, 부산항 전체에 대해서도 이상치를 감지할 수 있었다. 또한 그 이상치의 주요 원인에 대해서도 파악할 수 있었다.

이러한 이상감지 분석을 통해 부산항의 물동량 흐름 패턴을 분석하고 해당 패턴 상·하한을 벗어나는 이탈을 체크해 주어 과거 비슷한 흐름을 보일 때 즉각적 대응이 가능해진다. 그 원인에 대해서는 추가적인 인터뷰 및 조사가 필요하지만 트렌드를 분석하고 그 이상 이탈에 대해서 자동으로 체크해 주는 시스템이 있다면 이상 감지에 대한 비용 및 시간을 절감할 수 있고 보다 빠른 대응이 가능하게 될 것이다.

5.2 연구의 한계점과 과제

5.2.1 한계점

많은 연구자들이 시계열을 두 가지로 쪼개어 규칙성을 가지는 패턴과 불규칙성을 가지는 패턴의 결합이라고 이야기하였다. 시계열 모형은 전통적으로 이 두 가지로 개발되었는데 통계 정보란 대상의 본질 및 관계에 변함이 없는 경우에 대한 정보를 제공하는 것이다. 그런데 이러한 시계열 자료가 시간이 지남에 따라 그 패턴이 일정하지 않다면 그 통계 정보는 효력이 떨어지게 된다. 부산항 환적 물동량 이상 감지를 할 때 ARIMA와 이동평균법(MA)등의 시계열분석을 하게 되는데 이것은 그동안 일어났던 과거의 추이 분석 패턴들이 일정하지 않게 되면 그 설명력이 떨어지게 된다.

또한, 시계열 분석은 과거의 데이터를 통해 미래를 예측하게 되는데 시계열은 과거의 정보를 통해 얻어진 결과이다. 과거를 통해 미래를 알 수 있다는 대전제를

부정하게 되면 시계열이 미래를 예측하는데 바람직하지 않게 된다.

시계열이란 시간에 따라 결과가 중첩되어 상호작용하게 되며 시간이 지날수록 오차가 중첩되게 되어 분산이 크게 증가하게 되고 예측력이 사라지게 된다.

그러나, 시계열은 과거로부터 현재까지 정보에 대해서는 강력하다. 앞에서 실시한 부산항 전체 환적화물 물동량 이상감지 방법(ARIMA 방법)과 국가, 포트, 선사, 터미널별 이상 감지 방법(이동평균법, MA)은 과거 데이터에 대한 분석 결과를 가지고 초근접한 미래에 대한 대응이 가능하게 해 주기 때문에 일반적인 시계열의 한계점을 극복할 수 있게 된다. 또한, 부산항 전체 물동량 이상치 감지 뿐만 아니라, 국가, 포트, 터미널, 선사 등 개별적인 이상 감지 분석도 해 주기 때문에 이상 감지 원인을 분석하는데 더 효과적일 수 있다. 그러나, 과거 데이터 자체에 2009년 미국발 서브 프라임, 2016년 한진해운 사태와 같은 10년에 한번 있을까 말까 하는 초유의 사태까지 포함하여 분석하고 있어 해당 모형으로 미래를 예측하게 될 때 부정적인 오차가 생길 수 있다.

5.2.2 향후 연구 과제

해운시장 장기침체와 수급 상황 악화로 개별 선사들은 노선의 효율화를 통한 수익성 향상을 위해 선사의 전략적 인수 및 전략적 제휴 방법을 선택하고 있다. 이러한 얼라이언스는 급격하게 변화하는 시황에 대응하고 영업 손실을 최소화하는 방법 중 하나로 각광받고 있는 실정이다. 매년 1, 2회 정도의 얼라이언스 개편으로 선사들은 선대를 공유하여 윈윈 정책을 펼쳐 왔다. 이러한 얼라이언스 변화가 있을 때 해당 분석 모형에 얼라이언스 이벤트를 반영하는 설명변수를 추가하여 시계열 모형으로 이상감지 분석을 행하면, 얼라이언스 변동에 따른 평가가 가능하여 미래 대응에 도움이 될 것이다.

또한, 특정 부정적 이벤트를 제외하고 이상 변화 감지 분석을 행할 필요가 있

다. 2008년 9월 미국발 금융 위기나 2016년 9월 한진 해운 사태와 같은 매우 특수한 이벤트가 발생했을 때, 이러한 자료를 반영하여 모형을 분석하게 되면 향후 이상 감지 관리선이 왜곡될 가능성이 높다. 따라서 이러한 이상 상태 이벤트의 영향을 제거한 시계열 모형을 다시 수립하여 분석한다면 그런 문제를 보완할 수 있을 것이다.



참고 문헌

<국내 문헌>

- 천민경(2014), 빅데이터 역량 평가를 위한 참조모델 및 수준진단시스템 개발, 한양대
- 신현신(2016), 빅데이터 기반의 재난정보관리 방안, 서울시립대 석사학위논문
- 박두순(2014), 빅데이터 컴퓨팅 기술, 한빛아카데미
- 한국정보화진흥원(2017), 데이터 분석 전문가 가이드
- 이순교(2016), 빅데이터 시대에서의 스마트물류빅데이터 활용 사례 및 동향 고려대
- 빅데이터 센터(2016), 2015년 빅데이터시장 현황조사, 미래창조과학부, 한국정보화진흥원
- 이승희,김광일,박근철(2014), VTS BIG DATA를 활용한 해상교통 관제항로 패턴 분석,
한국해양항만학회
- 이재득(2013), 승법계절_ARIMA_모형에_의한_부산항_컨테이너_물동량_추정과_예측_해양대
- 이지홍 (2017), 빅데이터 분석을 통한 물동량 분석 예측 모델에 관한 연구 중앙대
- 강개선(2013) 부산항에서의 중국 환적화물 물동량 예측 부경대
- 김정훈(2008), 시계열 모형을 이용한 부산 북항의 물동량 예측, '한국항만경제학회지',
제24권 제2호, 2008, 1-17
- 송문섭, 허문열,(2002), 『수리통계학』, 박영사
- 김재경(2015), 혼합 극단분포를 이용한 이상점 탐색 연구 ,송실대
- 나은희(2016), 대용량 데이터 분석을 위한 맵리듀스 기반의 이상치 탐지 연구, 중앙대
- 하지현(2015) 이상치 탐지를 위한 강선한 알고리즘의 개발, 성균관대
- 김정훈(2008), 시계열 모형을 이용한 부산 북항의 물동량 예측, '한국항만경제학회지',
제24권 제2호
- 모수원 김창범(2003) 해상물동량 추정과 예측, 해운물류 연구, 제 37호, 1-18
- 양향진(2006), 중력모형을 이용한 부산항의 해상물동량의 입출항 패턴 분석, 한국항만
경제학회지, 제22권 제3호, 2006, 77-94
- 울산항만공사 웹진 강석환 (2017), 거스를 수 없는 4차산업혁명, 변화의 물결에 올라타라, SK해운

<외국 문헌>

Gartner (2011), Big Data Analytics, Gartner Group

Gartner (2011), Pattern-Based Strategy : Getting Value form Big Data, Gartner Group

Gartner (2011), How to plan Participate and Prosper in the Data, Economy, Gartner Group

McKinsey (2011), Big Data : The Next Frontier for Innovation, Competition, and Productivity,
McKinsey & Company

Cisco VNI Mobile (2014)

IDC & EMC Digital Universe Study 2011

<관련 사이트>

https://sustainability.ups.com/media/UPS_ORION_2016.pdf.

<http://retailnext.net/en/home/>.

<https://www.youtube.com/watch?v=KkkvOT4G6Ro>.

포트미스 홈페이지 <http://portmis.go.kr>

TUAS 항만 홈페이지 : http://www.enevproperty.com/?page_id=685

Bestfact, 'Smart Logistics at Hamburg Port'

http://www.bestfact.net/wp-content/uploads/2014/02/Bestfact_Quick-info_efreight_3-105_HPA-SmartPortLogistics.pdf

Box-Jenkins ARIMA모형에 의한 KOSPI와 Historical Volatility 이슈

출처 : [네이버 지식백과] 제4차 산업혁명 [The Fourth Industrial Revolution, 第4次産業革命

www.cs.uregina.ca/~dbd/831/notes/kdd.html

www.ibm.com

<https://www.techopedia.com/definition/25827/knowledge-discovery-in-databases-kdd>