

빅데이터를 활용한 해운시장분석 연구 - 빅데이터 파악 및 확보방안

A Study on Shipping Market Analysis
Using Big Data Analytics
- The Identification of Big Data

2017. 12.

윤희성·고병욱·박용안·박혜리·전형진·최석우·한성일



한국해양수산개발원
KOREA MARITIME INSTITUTE

보고서 집필 내역

〈연구책임자〉

윤 희 성 : 제1장, 제4장 1절, 2절, 4절, 제5장

〈연구진〉

고 병 욱 : 제2장 2절

박 용 안 : 제4장 3절

박 혜 리 : 제3장 1절, 2절(외부연구진 참여)

전 형 진 : 제4장 3절

최 석 우 : 제2장 1절, 제3장 3절

한 성 일 : 제2장 2절, 3절

산·학·연·정 연구자문위원

강 구 태 (위네텍 대표)

김 영 선 (현대해양서비스 팀장)

김 재 훈 (한국해양대학교 교수)

배 혜 림 (부산대학교 교수)

유 성 진 (한국해양대학교 교수)

* 순서는 산·학·연·정 순임

연구감리자

임 종 관 (한국해양대학교 초빙교수)

해운업에 있어 시장분석의 중요성은 아무리 강조해도 지나침이 없다. 시장 분석은 의사결정의 질을 높이는 기반이 됨과 동시에 해운업에 있어서 중요하게 관리해야 할 리스크 즉, 시장리스크와 신용리스크 관리의 기초가 되기 때문이다.

시장분석이 사업의 기초 역량으로서 중요한 역할을 함에도 불구하고 이 분야에 대한 연구는 학문적으로나 실무적으로나 제한적으로 이루어진 것이 사실이다. 이는 M. Stopford가 저서에서 언급하였듯이 지금까지의 시장분석이나 예측이 큰 성과를 내지 못한 데 기인한 것으로 생각된다.

이 연구는 빅데이터 분석이라는 새로운 방법을 해운시장분석에 도입하는 3단계 연구 중 1단계에 해당하는 ‘빅데이터의 파악 및 확보방안’에 대한 연구이다. 전통적인 시장분석이 한계를 보여 온 상황에서 빅데이터분석 방법론이 주목을 받는 데는 몇 가지 이유가 있다. 첫째, 전산처리 기술의 발달로 이전에는 상상하지 못했던 대용량의 데이터 처리가 가능해졌다. 둘째, 4차 산업혁명으로 사물인터넷을 통한 원시자료의 실시간 수집이 가능해졌다. 셋째, 기계학습기반의 분석방법이 도입됨으로써 판단 및 예측 분야에서 전통적인 분석기법을 능가하는 성과의 창출이 가능함이 밝혀지고 있다.

해운업에 대한 빅데이터 분석의 적용은 금융이나 원자재 분야에 비해 뒤처져있는 것이 현실이다. 설문조사에서 드러났듯이 많은 해운기업들이 시장의 분석과 예측에 과학적인 방법론을 적용하지 못하고 있으며 빅데이터 분석에 대해서도 아직 활용가능성이나 잠재된 역량에 대해 충분히 인지하지 못하고 있다.

해운업의 시장분석은 비단 해운업만의 문제는 아니다. 산업정책을 펴는 정부, 선박금융을 담당하는 금융기관, 해운을 수요로 하는 조선업, 해운 서비스를 공급받는 화주 등 많은 전후방 연관 산업에 큰 영향을 미치는 중대한 문제이다. 이러한 중요성에도 이 문제는 개별 해운기업의 차원에서 해결하는 데 많은 어려움을 겪어 왔다. 대기업은 조직운영의 연속성 문제로, 중소기업은 조직역량의 한계로 우리 해운의 위상에 걸맞는 역량을 키워오지 못했으며, 이러한 현상은 해운 불황기에 직면한 지금이나 앞으로 회복되는 단계에서도 지속될 것으로 예상된다. 이에 우리 연구원에서는 이러한 분석역량을 공적역량으로 확보하여 우리 기업에 제공함으로써 역량의 연속성을 확보하고 우리 해운업 및 전후방 연관산업의 국제경쟁력을 제고하고자 한다.

3단계에 걸친 이 연구를 완성하기 위해서는 해운, 금융, 전산, 해사기술 등

다양한 역량이 융합되어야 하며 데이터의 확보 및 공유를 위한 많은 정책적, 법률적 변화가 수반되어야 한다. 후속되는 연구가 데이터 플랫폼의 연구, 제도에 대한 연구 등으로 확대되어 빠른 시일 내에 우리 해운·조선업의 경쟁력 제고와 상생에 기여할 것을 기대해 본다.

이 연구는 빅데이터에 대한 연구이지만 엄밀한 의미에서의 빅데이터 즉, 대용량(volume), 빠른 처리 속도(velocity), 형태의 다양성(variety)이라는 조건을 충족시키는 데이터에 국한하지 않았다. 해운시장의 분석 단계에서는 이러한 조건에 부합되는 빅데이터 뿐만 아니라 전통적인 데이터도 구분 없이 사용될 것이기 때문이다. 따라서 연구의 범위에 자료 원천별로 광범위하게 흩어져 있는 전통적 데이터를 포함시켰으며 특히 상업적으로 제공되는 데이터의 구조를 면밀하게 파악함으로써 향후 분석단계의 활용을 촉진하는 데 의미를 두었다.

해운에 있어서도 빅데이터에 대한 크고 작은 연구차원의 접근이 있었지만 지금까지의 연구가 대부분 ‘분석’을 전제로 한 것이라기보다는 큰 플랫폼에 빅데이터를 어떻게 담을 것인가라는 내용이 주류를 이루었다. 그러한 측면에서 분석단계의 입력변수로서 데이터의 구조를 깊이 있게 다룬 이번의 연구는 이전의 연구와 차별성을 갖는다고 하겠다.

본 연구에는 연구책임자인 윤희성 부연구위원을 비롯하여, 전형진 해운시장 분석센터장, 한성일 통계분석실장, 박용안 연구위원, 고병욱 전문연구원, 최석우 전문연구원, 박혜리 전문연구원이 참여하였다. 그리고 한국해양대학교의 임종관 초빙교수와 유성진 교수, 김재훈 교수, 부산대학교의 배혜림 교수께서 보고서 품질 향상에 많은 도움을 주셨다.

또한 해운 업계에서는 현대상선, SM상선, 팬오션, 대한해운, 폴라리스 등 국내 우수 선사들의 임원들이 인터뷰와 설문조사를 통해서 좋은 의견을 개진해 주셨다. 이 연구에 협조해 주신 각 기관의 관련자들, 그리고 연구진들을 포함하여 연구보고서가 발간되기까지 많은 도움을 주신 모든 분들께 깊은 감사를 표한다.

2017년 12월
한국해양수산개발원
원 장 양 창 호



요 약 · i



제1장 · 1

제1절 연구의 배경 및 필요성	1
1. 해운시장분석의 중요성	1
2. 데이터 및 분석환경의 변화	5
3. 타 분야에서의 성과	6
제2절 연구의 목적	8
1. 단계별 연구내용	8
2. 연구의 목적	10
제3절 연구의 구성과 범위	11
1. 연구의 구성	11
2. 연구의 범위	12
제4절 연구 방법론	14



제2장 전통적 해운시장분석과 데이터 · 16

제1절 해운과 의사결정	16
1. 의사결정 이론의 발전	16
2. 해운기업의 의사결정	20
3. 해운의사결정의 산업간 연관구조	25

제2절 해운시장분석과 전통적 데이터	28
1. 정량적 시장분석	28
2. 정성적 시장분석	32
3. 전통적 데이터의 유형	33
4. 데이터의 활용 현황	36
제3절 전통적 시장분석의 현황과 한계	37
1. 시장분석 기법	38
2. 정량적 분석의 유용성	39
3. 분석 결과의 활용	40
4. 시장분석을 위한 조직	41
5. 전통적 시장분석의 한계	43



제3장 해운시장분석에 대한 새로운 접근법 -Big Data Analytics · 46

제1절 빅데이터 분석	46
1. 개요	46
2. 등장 배경	47
3. 빅데이터의 중요성 및 가치	49
제2절 빅데이터의 정의 및 분석기법	50
1. 빅데이터의 정의	50
2. 빅데이터의 특징	51
3. 빅데이터 분석단계	53
4. 빅데이터 분석기법	58
제3절 빅데이터 분석에 대한 연구	68
1. 해운에 대한 연구	68
2. 타 분야의 연구	72



제4장 해운빅데이터의 파악 및 확보 · 78

제1절 현황분석	78
1. 분석 방법	78
2. 현황	78
제2절 원천별 해운 빅데이터의 파악	81
1. 해운 관련 기관	81
2. 해운기업의 데이터	85
3. 데이터 공급 업체	86
4. 해운 빅데이터의 특징	91
제3절 해운 빅데이터의 활용사례	94
1. 해운기업 사례	94
2. 물류기업 사례	96
3. 항만 및 터미널 사례	99
4. 시사점	103
제4절 빅데이터의 확보방안	104
1. 사물인터넷 기반의 수집	104
2. 데이터 풀(pool)의 형성	106
3. 개방형 데이터에 연결	107
4. 수탁이나 공동연구를 통한 확보	109
5. 외부데이터의 구매	110
6. 웹탐색	111
7. 이해당사자 간의 교환	112
8. 추진방안	114



제5장 결론 및 정책제안 · 116

제1절 결론	116
1. 연구내용의 요약	116
2. 연구의 기대효과	119
제2절 정책제안	120
1. 100대 국정과제와 빅데이터를 활용한 해운시장분석 연구	120
2. 정책제안	122



참고문헌 · 125



부록 · 133

〈표 2-1〉 벌크해운과 컨테이너해운의 비교	21
〈표 2-2〉 중요도에 따른 해운의사결정의 우선순위	25
〈표 2-3〉 데이터의 공급원천	34
〈표 2-4〉 시장분석 담당 조직규모	42
〈표 3-1〉 빅데이터의 특징	52
〈표 3-2〉 데이터 전처리 기술 고려사항	55
〈표 3-3〉 변수 형태(독립/종속)별 모델링 방법	59
〈표 3-4〉 활용 목적(분류/예측)별 모델링 방법	60
〈표 3-5〉 R과 Python의 비교	67
〈표 3-6〉 데이터 종류와 수집 시 문제점	69
〈표 4-1〉 해양수산 빅데이터 추진계획 개요	82
〈표 4-2〉 IHS Markit의 서비스 목록	89
〈표 4-3〉 확보 방법별 특성	115

» 그림목차

〈그림 1-1〉 해운기업 도산의 구조.....	2
〈그림 1-2〉 원가수준과 손익의 실현.....	3
〈그림 1-3〉 3단계 연구의 개요.....	9
〈그림 1-4〉 1단계 연구의 개념적 구성.....	12
〈그림 1-5〉 연구의 범위.....	14
〈그림 2-1〉 의사결정지원시스템(DSS)의 발전단계.....	18
〈그림 2-2〉 의사결정지원시스템(DSS)의 구조.....	20
〈그림 2-3〉 해운업의 트레이딩 구조.....	24
〈그림 2-4〉 해운 의사결정의 연관구조.....	26
〈그림 2-5〉 데이터 공급원천별 점수합계.....	35
〈그림 2-6〉 데이터 활용현황 점수합계.....	37
〈그림 2-7〉 시장분석기법의 활용도 순위.....	39
〈그림 3-1〉 데이터의 분포도.....	46
〈그림 3-2〉 빅데이터의 유형 및 규모.....	48
〈그림 3-3〉 가트너그룹의 ‘The Big Data Value Model’.....	51
〈그림 3-4〉 빅데이터의 4가지 차원.....	51
〈그림 3-5〉 빅데이터의 표현기술.....	53
〈그림 3-6〉 빅데이터의 시각화 사례.....	57
〈그림 3-7〉 SEMMA 방법론을 활용한 데이터 마이닝 절차.....	61
〈그림 3-8〉 프로세스 마이닝의 연구 분야.....	62
〈그림 3-9〉 Text Mining Areas.....	63
〈그림 3-10〉 GPS신호의 분할.....	65
〈그림 3-11〉 2016년 빅데이터 분석 톨 사용설문 조사 결과.....	66
〈그림 3-12〉 지리적 시각화 예시(네덜란드 연안).....	70
〈그림 3-13〉 데이터 플로우차트.....	71
〈그림 3-14〉 공급망 관리(SCM) 모형 예시.....	73
〈그림 3-15〉 이탈리아 은행의 네트워크 모형.....	76
〈그림 3-16〉 예측모형의 시간이동(time shift) 개념도.....	77

〈그림 4-1〉 SP-IDC 시스템 개념도	83
〈그림 4-2〉 선박회사의 전산시스템 운영개요도	113
〈그림 5-1〉 3단계 연구의 기여	120

요약

빅데이터를 활용한 해운시장분석 연구 - 빅데이터 파악 및 확보방안

1. 연구의 목적

- 빅데이터를 활용한 해운시장분석 연구의 첫 단계에 해당되는 이 연구는, 분석모델 연구 및 시스템 구축을 위한 기초 작업에 해당되며, 연구목적은 아래의 두 가지로 구분됨
- 첫 번째 목적은, 데이터를 활용한 분석이 해운기업의 의사결정에 어떻게 활용되는지 그 현황을 명확히 파악함으로써 향후의 분석 및 시스템 구축 단계의 목적적합성을 제고하는 것임. 두 번째 목적은, 도처에 산재되어 있는 해운관련 데이터의 소재와 그 구조를 파악함으로써 분석단계의 입력변수를 체계화하는 것임

2. 연구의 방법 및 특징

1) 연구방법

〈표 요약-1〉 본 연구 방법의 특징

특징	주요 내용	자료수집	방법론 선택 이유
기초 분석	- 의사결정 이론의 발전 - 해운업의 의사결정 - 전통적 데이터와 시장분석 - 빅데이터 분석기법 및 사례	- 관련 문헌조사 - 국내외 사례조사	- 데이터 분석의 목적적합성 확보를 위한 의사결정 이론 및 실제 파악 - 전통적 데이터분석의 한계 및 개선방향 도출 - 사례연구를 통한 빅데이터 연구의 방향성 설정
설문 조사	- 해운 의사결정 - 해운시장분석의 현황 - 빅데이터 분석에 대한 인지도	- 11개 해운기업의 응답자 55명	- 해운 의사결정의 우선순위와 준거 파악 - 해운시장분석의 현황 파악 - 빅데이터 분석에 대한 인지도 파악
국내외 출장 조사	- 현행 시장분석의 현황 - 한계와 업계의 요구	- 국내 선사의 주요임원 면담	- 경영자 관점에서의 시장분석 필요성과 한계 확인
전문가 자문	- 빅데이터 분석기법 - 빅데이터와 기계학습기반 분석의 활용사례	- 분야별 전문가 자문 및 의견 청취	- 빠르게 발전하는 분석기법을 이해하고 적용사례를 확인함으로써 활용을 전제로 한 데이터 확보 가능

2) 연구의 특징

- 기존의 빅데이터 연구가 빅데이터의 활용방안보다는 플랫폼 구축방안을 주로 다루고 있었는데 반해 본 연구는 해운시장분석이라는 구체적인 목적을 설정하고 그 기초가 되는 데이터 파악 및 확보방안을 제시한 연구임
- 설문조사를 통해 해운업 의사결정의 우선순위와 의사결정 준거를 파악하고 전통적 데이터의 활용현황, 기존의 시장분석 현황을 파악함
- 빅데이터의 정의와 빅데이터 분석에 사용되는 기법들을 광범위하게 소개함으로써 향후 활용 측면에서의 이해를 도모함
- 해운 빅데이터를 원천별로 파악함으로써 향후 입력변수로 활용할 가능성을 높임
- 빅데이터의 확보방안을 웹탐색을 통한 수집에서 공동연구를 통한 확보까지 포괄적으로 다룸으로써 데이터 확보의 현실성을 확인함
- 정부 100대 국정과제의 관점에서 빅데이터 분석의 의미를 조명하고 빅데이터 분석의 지원에 필요한 정책의 방향성을 제시함

3. 연구 결과

1) 연구 결과 요약

- 해운업 의사결정의 반복된 실패는 산업의 위기를 초래하는 중요한 문제이며 의사결정은 정보에 기반을 둔 의사결정이 될 때 합리성과 질을 확보할 수 있음.
- 설문조사 결과, 해운기업의 가장 중요한 의사결정은 i) 사업구조의 결정과 ii) 선박의 확보/판매시점의 결정 및 계약기간의 선택으로 파악되었음
- 시장분석에 사용되는 데이터는 Clarkson에서 제공하는 시황보고서 및 시계열 자료에 대한 의존도가 큰 것으로 나타났음
- 데이터의 활용에 있어서는 50%가 넘는 응답자가 내부 분석보고서를 1순위로 꼽아 외부의 분석보고서를 그대로 의사결정에 반영하기 보다는 조합과 해석의 과정을 거쳐 활용하고 있는 것으로 밝혀짐

- 시장분석의 기법으로는 정량적 분석의 비율이 정성적 분석보다 크게 높아 시계열분석, 기술적 분석, 기술통계분석 등이 활발하게 시도되고 있음을 알 수 있음
- 시장분석 조직은 5명 이내의 소규모 조직으로 운영되는 경우가 많았으나 이 경우에도 영업기능의 일부로서 시장분석이 행해지는 것으로 파악되었음
- 빅데이터에 대한 인지도는 중간 수준을 유지한 것으로 평가되나 활용도 측면에서는 극히 제한적인 것으로 파악되었음
- 국내에서 생성되는 해운관련 원시자료(primary data)는 항만운영정보시스템(Port-MIS)에서 생성되는 자료와 관세청에서 생성되는 수출입관련 통계자료 등 제한적인 것으로 확인되었음. 그나마도 국내에 국한되어 있어 국제적인 성격을 띠는 해운시장의 분석 측면의 활용도에는 제한이 있는 것으로 판단됨
- 해운기업에서는 선박의 운항과정 중에 선박의 장비를 통해 사물인터넷을 활용하여 운항관련 정보를 세부적으로 확보하고 있음. 이들 자료는 운항효율성의 개선 측면에서는 매우 유용한 자료이나 시장분석의 관점에서는 활용에 제한을 가짐
- 해운시장분석에 활용될 수 있는 빅데이터인 AIS자료와 광범위한 시계열자료는 IHS Markit, Clarkson 등 해외의 상업적 공급업체로부터 유료로 확보할 수 있음
- 해운 데이터의 특징을 요약하면 아래와 같음
 - 국내 원천데이터가 부족함
 - 전통적 데이터 위주임
 - 공급자가 중복으로 존재함
 - 상업적 공급자가 주도적인 역할을 하고 있음
 - 공유와 개방이 미흡함
- 사례연구에서는 국내외적으로 해운, 항만, 항만당국 등 다양한 주체들이 4차 산업혁명의 기술을 활용하여 데이터를 확보하고 기계학습기반의 분석을 시도하고 있는 것으로 확인됨
- 빅데이터의 확보 측면에서는 다양한 방법이 있을 수 있음. 이 연구에서는 가능한 방법들을 도출하고 각 방법의 특성을 아래와 같이 요약함

〈표 요약-2〉 빅데이터 확보의 방법별 특성

	확보가능성	효과성	경제성
사물인터넷기반의 수집	낮음	낮음	낮음
데이터 풀의 형성	중간	낮음	높음
개방형데이터에 연결	중간	중간	높음
수탁이나 공동연구를 통한 확보	중간	낮음	높음
외부데이터의 구매	높음	높음	낮음
웹탐색	중간	중간	중간
이해당사자 간의 교환	낮음	낮음	높음

2) 기대효과

- 빅데이터를 활용한 시장분석연구가 본 연구를 기반으로 완성될 경우, 해운기업의 경영실패를 예방하고 위기 시 영향을 최소화하는 효과를 거둘 수 있을 것으로 예상됨
- 또한 관련 산업인 조선업에 대해서도 위기에 선제적으로 대응할 수 있는 합리적인 근거를 제시할 수 있으며 해양금융기관의 신용관리를 강화하여 부실채권의 방지를 도모할 수 있음
- 정부는 선제적으로 정책을 시행하여 위기를 예방하고 위기 시 공적 위기대응비용을 크게 줄일 수 있을 것임

3) 정책제언

- 원천데이터의 확보 및 개방이 필요함. 특히 AIS데이터를 공공데이터로 확보하여 해사안전 및 상업적인 의사결정에 활용할 수 있도록 하는 방안이 강구되어야 함. 데이터의 개방측면에서도 관련제도와 법령을 조속히 정비해서 개방의 속도를 높여 나가는 것이 중요함
- 해운시장분석이 효과적으로 이루어지기 위해서는 다양한 역량의 융합화가 필수적임. 따라서 융합인력이 조속히 양성될 수 있는 기반이 마련되어야 함
- 국내 여건 상 해운기업이 지속성을 가지고 시장분석역량을 보유하기 어려운 구조이므로 시장분석역량을 공공역량화 하는 방안이 강구되어야 함. 한 국해양수산개발원의 경우 정책연구기관이라는 한계가 있으나 외연을 확장하여 민간의 수요에 대응하는 시장분석연구를 수행하는 것이 바람직할 것으로 판단됨

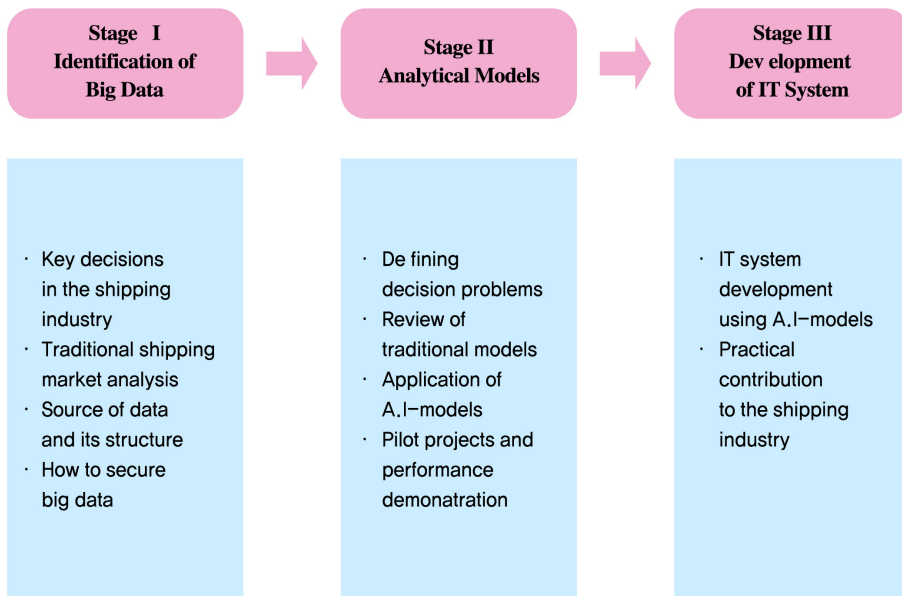
Executive Summary

A Study on Shipping Market Analysis Using Big Data Analytics

– The Identification of Big Data

1. Purposes

This research is part of a 3-stage big data analysis that has been applied to the shipping industry. The first stage is the identification of available big data that can be used to analyze the shipping market. The following stages are machine-learning-based analytical modelling and system development, respectively.



The eventual contribution of the research will materialize at the third stage, when the analytical system is developed and the output is visualized. The immediate purpose of this research is to serve the second and third stages of the research by identifying input variables of the analytical models.

The scope of this research was narrowed down to an analysis of the shipping market, which meant that the data had to be related to actual shipping decisions. Therefore, the decisions of shipping companies were analyzed before the data was identified.

The ultimate purposes/goals of the entire research can be summarized as follows:

- to enhance the competitiveness of the shipping industry,
- to prevent shipping and ship-building industry failure, and
- to minimize the impact when the industry faces a crisis.

2. Methods and Features

1) *Research methods*

This is a descriptive research focusing on the identification of data. Data-based decisions in the local shipping industry have not drawn much attention so far and shipping-related data has lacked a full-scale examination. The exploration itself did not require the application of specific research methods. A survey by means of a questionnaire and unstructured interviews, however, were employed in this research to figure out how traditional data and analytical methods are currently used in the shipping industry.

The questionnaire was designed to discover:

- decision priorities of the shipping industry,
- analytical methods to be used in current market analysis,

- the organization for market analysis,
- value and the use of market analysis,
- data used for the analysis, and
- the industry's understanding of big data analysis.

The questionnaires were distributed to 68 people in 11 shipping companies, the size of which varied from small to large entities. Fifty-five of the distributed questionnaires were collected, giving the researchers a response rate of 81 percent.

The identification of the big data required for shipping market analysis was carried out by web searching and expert interviews. The data was rather limited in scope and there was little to add to the knowledge of the researchers in terms of the data sources.

2) Features

This research was carried out with a focus on specific problems of shipping market analysis. Most reports on big data deal with the formation of a big data platform and ways to integrate big and/or traditional data. The identification of data in relation to actual decision-making for problems in the shipping industry is the main topic of this report.

The accumulation of data is meaningful only when the use of the data is taken into account. This research is significant in the sense that the data serves a specific purpose as part of a 3-stage system development plan. The plan is briefly displayed in Fig. 1.

Thirdly, the data was identified by both its source and its structure. The data structure of several important data such as the SIN of Clarkson Research and the IHS Markit was fully scrutinized and listed in the form of a separate booklet.

Another feature of this study is the survey on the current situation in the Korean shipping industry regarding decision making, decision criteria,

the use of data, analytical methods and research organization. In addition, the current understanding of big data analytics was examined to discover how local shipping companies are responding to fast-changing business environment.

The final feature is that the study is directly and closely related to the crisis that the industry is currently facing. The Korean shipping industry, as well as the shipbuilding and ship finance industries, have been suffering from long-lasting market depression since the financial crisis in 2008. This research can act as a detailed solution to these problems by enhancing the competitiveness of the Korean shipping industry.

3. Results

1) *Summary*

Decisions regarding business structure and purchase/sales of carrying capacity were revealed to be the most important decisions in the shipping industry. Both decisions need to be made rational by being supported by data analysis.

The decisions made by local shipping companies lack scientific foundation. Many shipping entities accept an amalgamation of external market reports—rather than scientific data analysis—as a decision criterion. Even though about 58% of the respondents considered time-series analysis to be the most important analytical method, little credit was given to the method and only 27% of the respondents recognized the usefulness of the method.

There could be a number of diverse reasons for the limitations of the quantitative method. One of the most plausible reasons is historic poor performance, as illustrated by M. Stopford.

The environment surrounding data analytics is experiencing a new

phase in the age of the so-called '4th industrial revolution'. The collection of big data using IoT technology is opening up a new horizon for data analytics and new A.I.-based tools are being recognized as a better-performing approach by demonstrating superior performance in the areas of finance and commodities.

This is the first part of the 3-stage research plan, of which the main idea is to examine how the shipping industry is reacting to the fast-changing data environment. This research consists of an understanding of the current market, the identification of data and approaches to data acquisition.

The survey showed that decisions made within the shipping industry have a weight on the formation of business structure, as well as the timing and quantity of freight trading. These decisions are considered to be more strategic and more closely related to market risk management than to decisions on operational efficiency.

The decisions are based on integrated 'internal' reports. This implied that shipping companies did not rely on external reports unless they were screened by internal experts. The internal reports were generally well accepted but the opinion of top-level management seemed to be influential and could change the views of the floor easily.

Time series analysis seemed to be the main analytical tool in use. The organization in charge of market analysis consisted of about 3-5 people, most of whom were not fully dedicated to market analysis.

An awareness of big data was not high, but 70% of the respondents held the view that big data analytics positively influences shipping business decisions. On the contrary, the collection, analysis and application of big data were found to be extremely limited.

In general, big data identification is carried out for primary data only. Very little primary data is produced locally. In the case of public data,

cargo movement data captured from Port-MIS and import/export statistics generated by the customs office are some limited examples. Cargo movement data from other public organizations are mostly secondary data reproduced by port authorities and the customs office.

Private companies are known to collect vessel operation data using IoT technology. Most of the data is gathered by the VDR and AMS installed onboard and periodically transmitted to shore-based computers using satellite communication system.

Commercial data agencies supply most of the data that can be utilized for market analysis. This data is mostly categorized as traditional data. The type of data that can be classified as big data is extremely limited. Automatic Identification System (AIS) data may be the only example in the category. The data in this report encompasses mostly traditional data such as those provided by Clarkson Research, IHS Markit and Baltic Exchange, as well as some AIS data supplied by IHS Markit and FleetMon.

There are diverse ways to secure big data:

- IoT-based capture
- Data pool
- Connection to open database
- Data sharing for commissioning and/or collaboration
- Web searching
- Data exchange
- Purchase

One of the most practical methods to secure data, from the perspective of a public research organization such as KMI, is through the purchase of data from commercial agents. Most of the time-series data and AIS data are available from them.

On the other hand, an economical and even creative way to acquire data can be from 3rd-party institutions such as IMO, UNCTAD and

classification societies that have a collection of data but lack the capabilities to analyze them. KMI could obtain the data free of charge in exchange for providing a data analysis service.

The value of the data increases as the number of observations grow. Companies and institutions could enjoy an improved analytical performance by putting their data in a pool. A research institute such as KMI could secure the data at no cost by offering an integration service.

The data can also be collected via web searching or through IoT-based capturing. This data can be unstructured and be classified as big data, which needs the application of distributed computing technology.

KMI at this stage does not generate any primary data. Hence, the exchange of data does not constitute a practical solution. The option can be considered when primary data is directly collected by KMI in the future.

2) Contributions to policy making

A few suggestions could be made to the government based on this research:

The market analysis needs to be provided to the industry as a 'public service'. The shipping industry seriously lacks the capability to analyze the market in a scientific manner and the development of a 3rd-party forecasting service does not seem to be feasible, at least in the near future. Under these circumstances, the only way for the government to improve the situation is to take action.

Investment in the development of human capital is essential. We are living in an age of convergence and the area of market analysis is no exception. Even though the industry has experts in each specific area, the expertise can easily be of little use unless they are integrated for effective problem-solving. The industry needs a group of experts who have good

domain knowledge and analytical techniques. The role of the government is essential in developing human resources.

More data needs to be open and made publicly available. A good analysis requires quality data and the data is often held in the public domain. While many advanced countries have a policy of opening their data to the public, Korea lags behind on such policies. For the data to be made publicly available, the government must provide initiatives for the amendment of relevant laws and regulations.

The industry is suffering from a chronically poor market and urgently needs to enhance its competitiveness. Quality market analysis is certainly at the core of this competitiveness. As mentioned above, the role of the government is essential in building up the industry's core competency. This report and successive research to follow will provide a good ground to improve the way in which shipping business is managed.

제1장

서론 <<

제1절 연구의 배경 및 필요성

의사결정의 합리화와 과학화라는 관점에서 해운시장분석은 기업의 성과와 가장 밀접하게 관련된 영역이라고 할 수 있다. 해운시장분석은 정량적 분석과 정성적 분석으로 나누어지는데 지금까지의 정량분석은 성과 측면에서 한계가 있어 실무에서 잘 활용되지 않았다.

최근 들어 전산처리 능력이 비약적으로 발전하고 데이터의 가용성이 증대됨에 따라 시장분석이 새로운 국면을 맞게 되었고, 학문적인 연구나 실무적인 적용 측면에서 앞서나가고 있는 금융이나 원자재 부분에서는 이러한 환경의 변화를 활용한 새로운 접근이 시도되고 있다. 특히 기계학습(machine learning) 기반의 분석도구들이 다양한 분야에 적용되어 좋은 성과로 주목을 받고 있는 것이 현실이다.

하지만 해운 영역에 이러한 접근이 이루어진 사례는 아주 드물어 연구의 필요성이 높다. 이 절에서는 해운시장에 있어 분석의 중요성, 전산처리를 둘러싼 환경의 변화 등 이 연구의 필요성과 배경을 다룬다.

1. 해운시장분석의 중요성

해운업을 영위하는 데 있어 해운시장분석이 왜 중요한지는 해운업 실패의 관점과 해운업 리스크관리의 관점으로 나누어 설명할 수 있다.

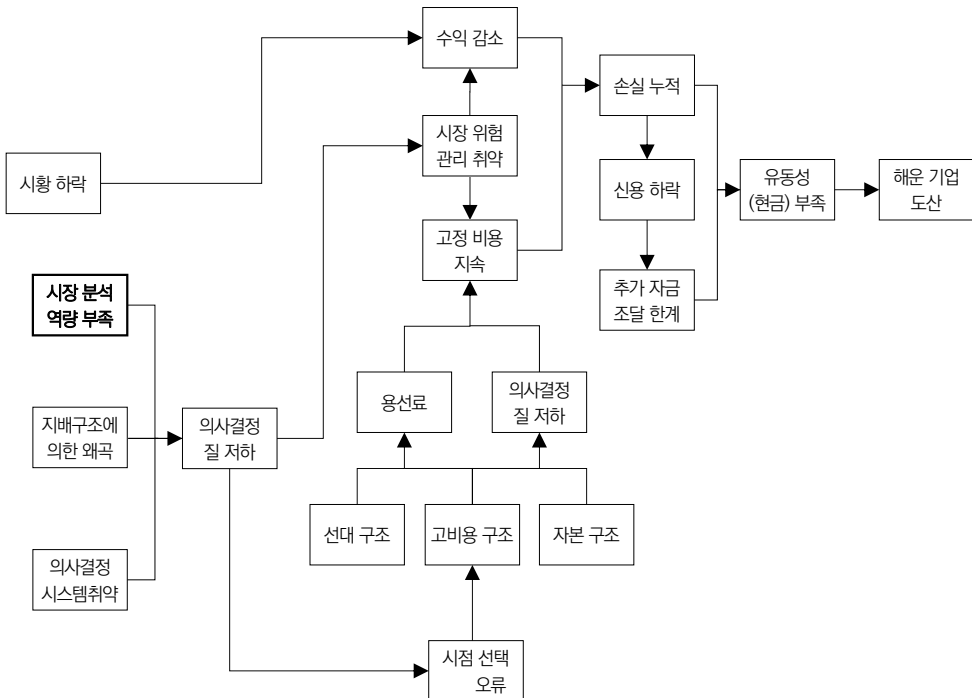
1) 해운업 실패의 예방

해운업은 대표적인 경기산업으로서 시간의 흐름에 따른 극심한 가격(운임)의 변동성이라는 특성을 가지고 있다. 우리나라의 해운업은 호황기에 규모를 확장

그 기간 중 누적 적자는 약 10조 원에 달한다.¹⁾

도산에 영향을 미치는 요인의 작용 구조를 나타낸 것이다.

〈그림 1-1〉 해운기업 도산의 구조

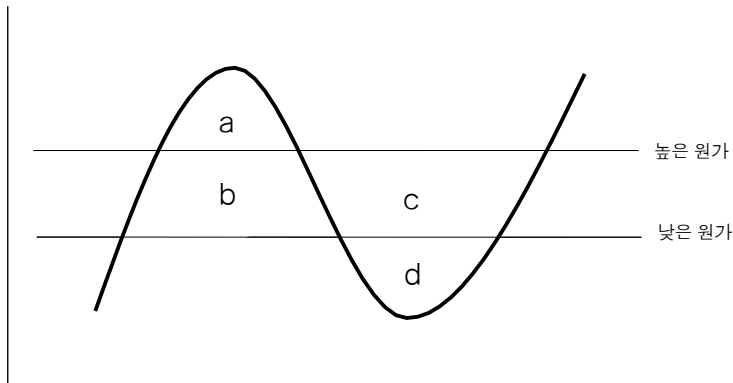


자료 : 저자 정리

1) 선주협회, 2017해양산업 CEO포럼 발표자료, 2017. 10.24

해운기업 도산의 구조를 보면 시장분석 역량이 해운기업의 실패에 미치는 영향을 구조적으로 파악할 수 있다. 시장분석 역량의 부족은 의사결정의 질을 저하시키는 주된 요인이며 의사결정의 질은 시장위험의 관리에 직접적인 영향을 미친다. 대표적인 예가 시황 고조기에 선대를 확보하여 순노출(net exposure)을 증가시키고 반대로 저조기에는 선대의 확보를 제한하는 의사결정을 함으로써 시황하락기에 손익의 변동폭을 확대하는 것이다. 시장분석의 또 다른 기능은 시점선택의 오류가능성을 줄이는 것이다. 체계적인 시장분석은 투자 또는 처분 시점의 선택에 대한 오류의 가능성을 낮추는 데 기여할 수 있다. 시점의 선택이 잘못되면 경쟁선사보다 높은 수준에 선대의 원가를 고정하기 때문에 불황에 견디는 힘이 현저하게 취약해진다. 그림 1-2는 원가수준이 기업의 손익에 미치는 영향을 도식화한 것이다. 원가가 높을 경우 이익은 a, 손실은 c+d가 되고 원가가 낮을 경우 이익은 a+b, 손실은 d가 되어 원가구조가 낮은 기업이 호황기에 더 큰 이익을 축적할 수 있고 불황기에도 더 큰 생존능력을 갖게 된다.

〈그림 1-2〉 원가수준과 손익의 실현



위에서 보듯이 시장분석능력은 기업의 원가경쟁력을 결정하는 핵심요소이며 위기 시 기업의 생존능력을 결정하는 중요한 요소로 작용한다. 시장분석은 데이터의 분석을 통해 이루어진다. 따라서 의사결정과 관련된 양질의 데이터를 확보하는 것은 과학적인 분석과 예측의 기초로서 무엇보다도 선행되어야 한다.

2) 관련 산업에 대한 영향의 파악

해운의 대표적인 전후방 연관 사업으로는 무역, 조선, 금융을 들 수 있다. 해운업에 있어 수익의 원천이 되는 해상운임이 화주에게는 물류비의 중요 구성요소가 된다. 과학적인 시장분석에 기초한 합리적인 시황의 예측은 선주와 화주 모두에게 장기적인 수익, 비용 안정화를 가져다주는 요인이 된다. 고시황기에 이어지는 하락에 대한 예측이나 저시황기에 뒤따르는 상승에 대한 예측에 신뢰성이 가해지면 선주나 화주를 막론하고 변동성을 축소하고 수익이나 원가 구조를 안정화시키는 데 이해기반을 같이 할 수 있어 장기계약을 통해 윈-윈 상황을 만들어 낼 수 있다.

해운은 조선의 수요산업이기 때문에 해운시장분석은 조선소 수요예측에 결정적인 영향을 미치는 요소가 된다. 우리나라 조선산업이 해운 경기의 부침에 따라 손익의 변동성을 키운 것은 분석에 기초한 사전적인 대응보다는 현상에 대한 대응에 머물렀던 것이 원인이다. 만일 해운시장에 대한 과학적인 분석으로 미래에 대한 가시성을 키운다면 사전적인 전략수립으로 손익을 안정화 하는 데 기여할 수 있을 것이다.

선박금융의 관점에서 볼 때 해운기업의 시장위험은 금융기관의 신용위험 분석의 기초가 된다. 금융기관의 기업에 대한 신용분석이 대부분 기업의 재무제표에 근거하여 사후적이고 정태적인 형태로 이루어지고 있지만 시장분석이 발전하면 사전적이고 동태적인 신용위험의 관리가 한층 용이해질 것이다.

3) 정부 정책판단의 기초

정부는 산업정책의 실행주체로서 산업의 동향을 정확히 파악하여 정책을 수립·실행할 필요가 있다. 이를 위해서는 시장의 수급상황과 운임시장의 추세를 과학적이고 합리적인 방법으로 분석하고 예측할 수 있어야 하는데 그 중에서도 조기경보체계를 운영하여 사전적인 정책수립으로 위기를 예방하고 위기 도래시의 영향을 축소하는 것은 중요한 역할이라고 할 것이다. 이러한 의미에서 데이터에 기초한 과학적인 분석역량의 확보는 정부의 정책역량과 직결된 중요한 과제이다.

2. 데이터 및 분석환경의 변화

데이터 처리와 분석을 둘러싼 환경은 지난 10년간 큰 변화를 겪어왔다. 이러한 추세는 분산처리 기술을 활용한 빅데이터의 처리, 사물인터넷(IoT)의 발달에 따른 새로운 형태의 데이터 활용, 기계학습(machine learning) 기반 분석기법의 발달로 요약될 수 있다.

이러한 환경의 변화는 해운시장분석에 대한 새로운 접근을 촉발하여 데이터의 파악과 분석에 대한 연구의 필요성을 부각시키고 있다.

1) 새로운 형태의 데이터

데이터 분석에 있어 도처에 산재되어 있는 기존의 데이터를 정확히 파악하고 이를 통합하여 분석에 활용하는 것도 중요한 문제이나 새로운 형태의 데이터가 다양한 방법으로 수집되거나 축적되어 분석에 활용될 수 있다는 사실에 초점을 맞출 필요가 있다.

빅데이터는 흔히 3V로 정의된다. 이는 데이터의 크기(volume), 축적과 처리의 속도(velocity) 그리고 형태의 다양성(variety) 측면에서 전통적 데이터의 범위를 뛰어 넘는 데이터로 빅데이터를 정의하는 것이다.

빅데이터는 주로 기계장치에 부착된 감지센서 또는 웹탐색을 통해 수집되며, 활용에 있어서는 해결하고자 하는 문제와 창의적으로 연결하는 것이 관건이 된다. 교통카드의 사용 즉, 특정 도시의 인구이동 기록이 부동산 상권분석이라는 문제의 해결에 활용되는 것을 예시로 들 수 있다.

해운에서는 선박의 장비들이 전자적으로 운영되고 통제됨으로써 자동적으로 생성되는 데이터가 있는데 이를 감지센서를 통해서 수집하는 것이 빅데이터의 성격을 가지고 있으며 전세계 선박의 이동경로를 파악하는 선박의 자동식별시스템(AIS: Automatic Identification System)이 생성하는 데이터도 빅데이터의 범주에 들어간다.

시장분석의 관점에서는 이러한 데이터를 상업적인 의사결정 문제에 연결시키는 것이 중요한 역량이 된다. 이를 위해서는 IT지식과 분석기법에 더하여 해운업에 대한 도메인지식(domain knowledge)²⁾을 확보하는 것이 중요하다.

2) 빅데이터 처리기술

개념적으로 빅데이터는 독립된 컴퓨터의 처리능력을 초과하는 분량의 데이터이기 때문에 처리기술의 핵심에 분산처리가 있다. 2005년에 Doug Cutting과 Mike Caferella에 의해 Aapche Hadoop이 만들어진 이래 빅데이터 처리기술은 비약적인 발전을 이루었다.

현재 해운에 있어 분산처리를 요하는 대규모의 데이터는 그다지 많지 않은 것으로 파악되었으나, 외견상 관련성이 거의 없는 데이터가 분석을 통하여 실질적으로 의사결정을 지원하는 경우가 많으므로 데이터에 대해서는 개방적인 접근을 유지할 필요가 있다.

3) 기계학습기반의 분석도구

데이터는 분석의 과정을 거쳐 사업상의 가치(business value)를 창출하게 된다. 전통적 데이터의 경우 주로 정량적인 시계열 데이터를 계량경제학적 기법이나 기술적 분석기법(technical analysis)을 통하여 분석하는 것이 일반적인데 해운에 있어서는 역사적 성과가 그다지 좋지 않아 활발하게 사용되지 못했다.

하지만 기계학습기반의 예측모델링(predictive modeling) 기법이 적용되면서 다양한 분석이 활용되고 있고 해운 이외의 다른 영역에서는 그 성과가 전통적인 방법보다 우위에 있다는 점이 드러나고 있다. 이러한 기법 상의 차이들도 해운영역에 대한 데이터와 분석에 대한 연구의 필요성을 일깨우고 있다. 구체적인 분석기법에 대해서는 3장에서 다루기로 한다.

3. 타 분야에서의 성과

기존의 계량경제학 방법론이 변수들 간 인과관계의 증명을 목적으로 정형적 모형 제시를 추구하였다면 최근의 빅데이터를 활용한 인공지능, 기계학습, 딥러닝

²⁾ Domain knowledge is valid knowledge used to refer to an area of human endeavour, an autonomous computer activity, or other specialized discipline. (Wikipedia),
<http://en.wikipedia.org/wiki/Domain-Knowledge>, 검색일 : 2017.12.1

등의 새로운 기법들은 변수들의 다차원 상관관계 파악을 통해 예측(forecasting)의 정확도를 높이고 있다. 단기 또는 중장기 시황 예측이 주요 분석 목적이 되는 금융 분야는 특히 이 새로운 기법들을 빠르게 받아들이고 있다. Shen(2012)³⁾은 주식시장 예측에 SVM(Support Vector Machine) 기반의 기계학습 알고리즘을 적용하여 타 분석법 대비 향상된 예측 정확도를 보였다. Li(2017)⁴⁾는 텍스트마이닝(text-mining) 기법인 감성분석(sentiment analysis)과 기존의 예측모형이 결합한 하이브리드 모형을 적용하여 서부텍사스산원유(WTI) 선물 가격의 추세를 예측하였으며, Wimmer(2018)⁵⁾는 최근의 인공지능 기법이 시장의 비효율성 측정, 신용점수 기준 설정, 파산가능성 예측 등 각종 금융투자 시나리오에 적용되어 예측 정확도를 향상시킨 사례를 요약 정리하였다.

학계의 연구결과들이 축적되면서 금융회사들은 빅데이터 분석을 업무에 본격적으로 활용하고 있다. 초기에는 소비동향에 민감한 신용카드회사와 보험회사를 위주로 빅데이터 활용이 활성화되었다면 이제는 시중은행들도 빅데이터 관련 조직을 대폭 강화하는 등 빅데이터를 은행 경영에 접목시키는 시도를 하고 있다. 정해용(2017)⁶⁾에 따르면 독일 도이치은행은 SNS 기반의 빅데이터 분석을 통해 고객의 신용도를 분석하고 있으며, 미국 웰스파고은행은 고객의 ATM 활용 패턴을 분석해 고객 개개인에 특화된 ATM 유저인터페이스를 제공하고 있다. 국내에서는 신한은행이 빅데이터센터를 통해 연립다세대주택 시세산정 시스템을 개발하였고, KB국민은행은 고객의 빅데이터를 분석해 맞춤형 마케팅이 가능하도록 하고 있다. 또 우리은행은 빅데이터 활용 리스크 분석모델과 고객수익률 관리 등의 서비스를 제공할 예정이다.

한편 배준호(2017)⁷⁾에 의하면 2017년 3월 세계 최대 자산운용사인 블랙록

3) Shen, S., H. Jiang and T. Zhang (2012), "Stock market forecasting using machine learning algorithm", Semantic Scholar.

4) Li, J., Z. Xu, H. Xu, L. Tang and L. Yu (2017), "Forecasting Oil Price Trends with Sentiment of Online News Articles", Asia-Pacific Journal of Operational Research, Vol. 34, No. 02.

5) Wimmer, H. and R. Rada (2018), "Applying Artificial Intelligence to Financial Investing", *Encyclopedia of Information Science and Technology*, ed. 4, Ch. 1.

6) 정해용, 금융 빅데이터 전쟁 - 국내 금융사, 빅데이터 활용 쟁결음(2017.06.13.),

http://biz.chosun.com/site/data/html_dir/2017/06/12/2017061201897.html, 검색일 : 2017. 12. 5

7) 배준호, 스타 머니매너지 시대가 끝나고 있다(2017.03.29.), <http://www.etoday.co.kr/news>

(Blackrock)은 주식투자 종목 선정에서 컴퓨터 모델에 중점을 둔 주식사업 개편안을 발표하였다. 블랙록의 새 전략은 미국 대형주 투자에서 인간이 시장을 이길 수 없다는 믿음에서 비롯된 것으로, 스타 펀드매니저가 중심이 된 능동형 투자 시대에서 빅데이터와 인공지능을 활용하여 알고리즘과 데이터에 바탕을 둔 시스템적인 투자로의 전환이 시작된 것으로 평가된다.

금융 분야에서 빅데이터와 인공지능을 활용한 다양한 시도와 가시적 성과들은 해운 분야에 시사하는 바가 크다. 전통적 시장분석 방법론의 여러 한계를 극복하고 빠르게 변화하는 환경에 유연하게 대처하기 위해 해운 분야 역시 빅데이터분석 방법론에 주목하고 실제 활용을 위한 적합성과 타당성 연구가 필요하다.

제2절 연구의 목적

연구의 목적을 기술하기에 앞서 이 연구를 포함한 전체적인 연구의 단계를 정리할 필요가 있다. 데이터의 확보방안에 대한 연구에 이어 분석모델에 대한 연구가 이루어져야 하며, 궁극적으로는 분석모델의 구현을 통해 실질적인 기여가치가 창출되어야 하기 때문이다. 연구의 목적은 1단계에 해당하는 이 연구의 목적과 3단계 연구 전체의 목적으로 구분하여 기술한다.

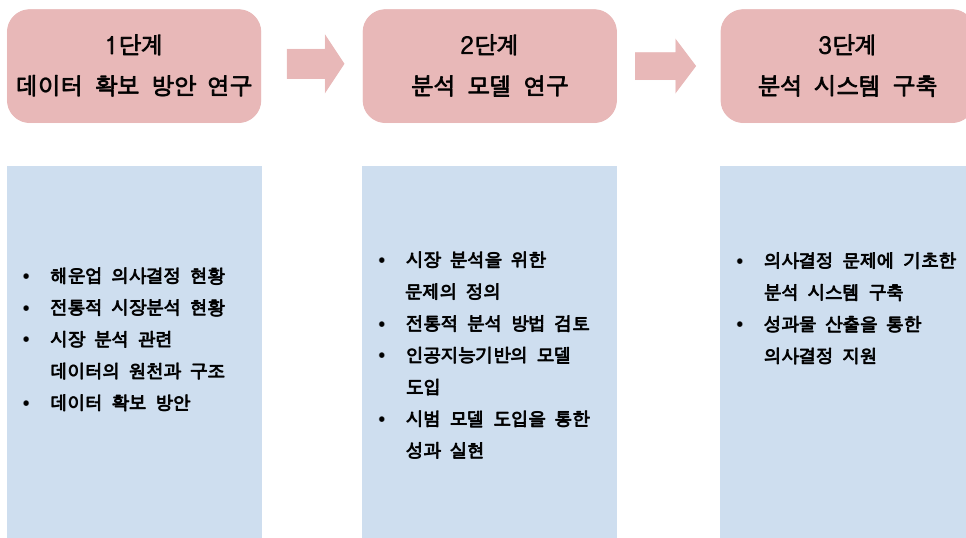
1. 단계별 연구내용

빅데이터를 활용한 시장분석 연구는 3단계로 이루어진다. 1단계는 데이터의 원천과 구조를 파악하고 확보방안을 정리하는 것을 내용으로 한다. 파악하는 데이터의 범위를 정하기 위해서는 해운의 의사결정과 자료의 관련성이 우선 파악되어야 한다. 따라서 해운의 의사결정을 유형별로 정리하고 그러한 의사결정에 지금까지 어떠한 자료들이 활용되었는지 현황을 파악하는 것이 선행된다.

2단계는 분석모델에 대한 연구이다. 해운시장의 분석에는 시계열분석이나 기술적 분석 등 일반적인 정량적 분석기법이 활용될 수 있으나 실무적으로 그러한 정량적 분석기법들이 비중을 가지고 다루어지지 않았다. 최근에 대두되고 있는 기계학습기반의 분석 및 예측모델들은 재무나 원자재 등의 영역에서 전통적인 예측기법의 성과를 능가하는 것이 입증되고 있으므로 이를 중심으로 모델연구를 진행한다. 모델의 연구과정에서 소수의 시범모델(pilot model)을 구축하여 성과를 시험함으로써 3단계 연구를 견인한다. 2단계의 연구는 학·연 공동으로 진행되며 이러한 연구조직 구성은 3단계까지 지속된다.

3단계는 1단계와 2단계의 연구를 종합하여 실제 모델을 구축함으로써 정보의 수요자에게 유의미한 정보를 제공하고 기여가치를 창출하는 단계이다. 이 단계에서는 의사결정요구에 기초한 문제의 설정이 중요하므로 산·학·연 공동으로 진행된다.

〈그림 1-3〉 3단계 연구의 개요



2. 연구의 목적

1) 이 연구의 목적

빅데이터를 활용한 시장분석 연구의 1단계에 해당되는 이 연구는, 2, 3단계 연구를 위한 기초 작업에 해당되며, 연구목적은 아래의 두 가지로 구분된다.

첫 번째 목적은, 데이터를 활용한 분석이 해운기업의 의사결정에 어떻게 활용되는지 그 현황을 명확히 파악함으로써 향후의 분석 및 시스템 구축단계의 목적 적합성을 제고하는 것이다.

두 번째 목적은, 도처에 산재되어 있는 해운관련 데이터의 소재와 그 구조를 파악함으로써 분석단계의 입력변수를 체계화하는 것이다.

해운관련 데이터는 전통적인 데이터의 경우에도 기업, 기관, 상업적 공급업자 등 다양한 주체가 보유하고 있어 그 실체가 명확하지 않다. 빅데이터의 성격을 가진 데이터도 그 형태나 확보방법에 대한 선행연구가 별로 없다. 지금까지의 빅데이터 연구가 분석을 위한 입력변수로서의 데이터에 초점을 맞추기 보다는 빅데이터 플랫폼에 대한 연구 형태로 이루어진 것도 선행연구가 부족한 이유 중의 하나이다.

2) 3단계 연구의 최종목적

1단계 연구로서, 해운관련 빅데이터 자료에 대한 연구를 수행하는 과정에서 3단계 연구의 최종목적을 기술하는 것이 다소 성급해 보일 수 있으나, 1단계 연구 목표는 최종적인 성과를 창출하기 위한 ‘과정’상의 목적이므로 위의 단계별 연구내용에서 기술된 3단계가 완성되었을 경우를 가정한 목적을 명확하게 적시할 필요가 있다.

빅데이터 해운시장분석을 통한 3단계 연구의 궁극적인 목적은, 해운시장분석에 새로운 과학적 기법을 적용함으로써 의사결정의 신뢰도를 제고하고 해운업의 경쟁력을 강화하는 것이다. 이를 통하여 해운기업의 경영 실패를 예방하고, 위기 도래 시에도 그 영향을 최소화할 수 있을 것이다.

3단계 연구의 최종적인 결과는 해운업뿐만 아니라 관련 산업인 조선업, 수출입의 주체인 화주, 선박금융을 담당하는 금융기관, 정책당국인 정부에 의해 고르게 활용될 수 있다.

제3절 연구의 구성과 범위

1. 연구의 구성

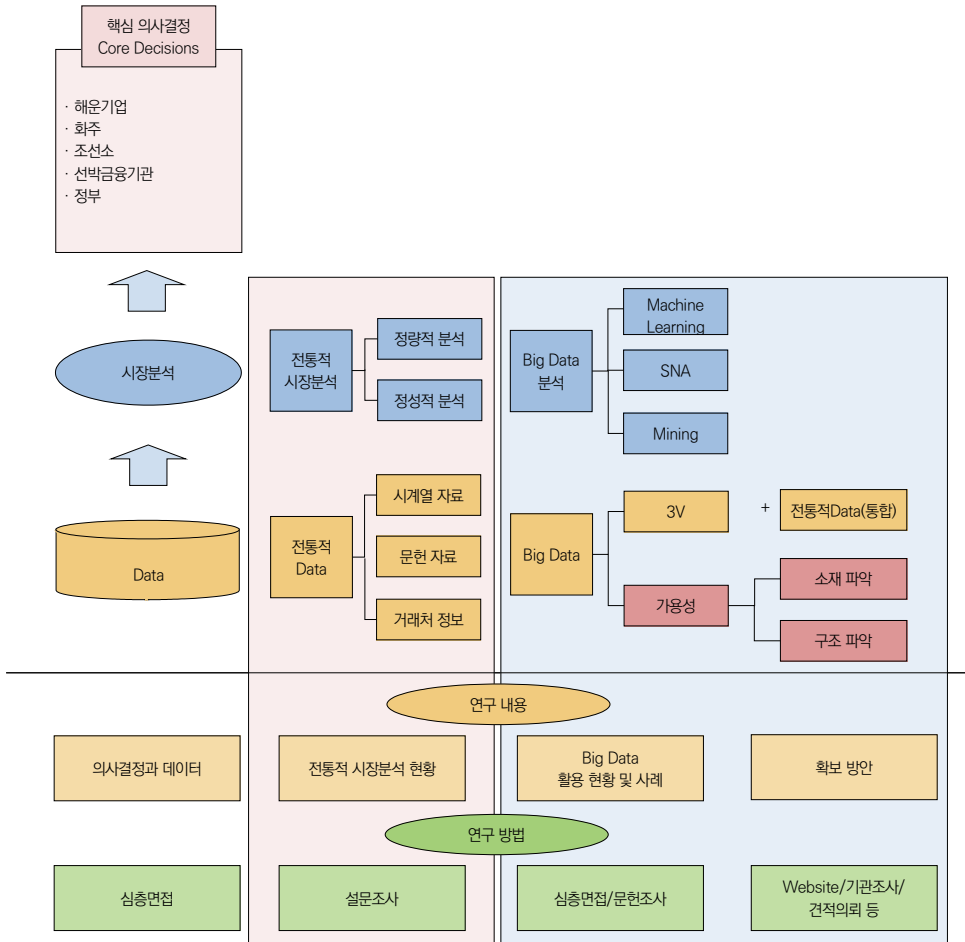
이 연구는 해운기업에서 어떠한 전략적인 의사결정이 이루어지고 있는지, 그리고 그 의사결정을 뒷받침하기 위해서 어떠한 정보들이 어떻게 분석되어 활용되고 있는지를 파악하는 것으로 시작된다. 해운업은 다양한 형태를 띠고 있고 사업의 속성도 크게 다르므로 사업의 유형에 따라 의사결정이 어떠한 차이를 갖는지에 대해서 우선 살펴볼 필요가 있다. 여기서는 컨테이너와 벌크를 구분하고 의사결정 속성이 유사한 선주업을 벌크선과 같은 범주에 넣어서 사업모델을 두 가지로 대별하였다. 다음 단계는 기존의 해운시장분석기법과 분석에 활용된 전통적인 데이터를 파악하는 것이다. 해운시장분석에 있어 정량분석은 매우 제한적으로 활용되어 왔으므로 학문적으로 연구된 분석기법과 실제 활용도를 분리해서 다룰 필요가 있다.

3장에서는 새로운 방법에 해당하는 빅데이터의 기본 개념과 기계학습 기반의 데이터 분석방법에 대해서 포괄적으로 소개한다. 이 부분은 2단계 연구에서 활용될 방법론을 소개하는 측면에서 의미가 있다.

4장에서는 해운시장분석에 활용되는 빅데이터의 소재를 원천별로 파악하고 그 구조를 파악한다. 구조는 향후 분석 시 입력변수로 활용될 수 있도록 데이터의 시점과 종점, 빈도, 항목 등을 구체적으로 파악한다. 빅데이터를 시장분석에 활용하기 위해서 어떻게 확보할 것인가를 다루기 전에 해운 관련 빅데이터가 어떻게 활용되었는지 파악되어야 한다. 이어서 최근에 빅데이터 분석이 해운업에 있어 활용된 사례를 해운업과 관련 산업으로 구분하여 소개한다. 원천별로 파악된 해운관련 빅데이터를 확보하는 방안에 대해서 이 장의 후반부에서는 산재되어 있는 기존의 전통적 데이터를 연결 또는 구매하고 통합하는 방법과 이전에 활용되지 못했던 새로운 형태의 데이터를 수집하는 방안에 대해 다룬다.

5장에서는 연구를 요약하고 정책에 반영할 사안을 도출하여 제시하는 것으로 마무리한다. 연구의 개념적 구성을 그림으로 표시하면 아래와 같다.

〈그림 1-4〉 1단계 연구의 개념적 구성



2. 연구의 범위

빅데이터 분석은 매우 광범위하기 때문에 그 외연을 확장할 경우 효과적인 연구가 되기 어렵다. 따라서 이 연구에서는 연구의 범위를 아래와 같이 한정하여 목적적합성과 실현가능성을 제고하였다.

첫째, 제2절 1항에서 소개한 3단계의 빅데이터 연구 중 1단계에 해당하는 빅데이터 확보로 한정하였다. 물론 이를 위해서는 기존의 시장분석이나 전통적인

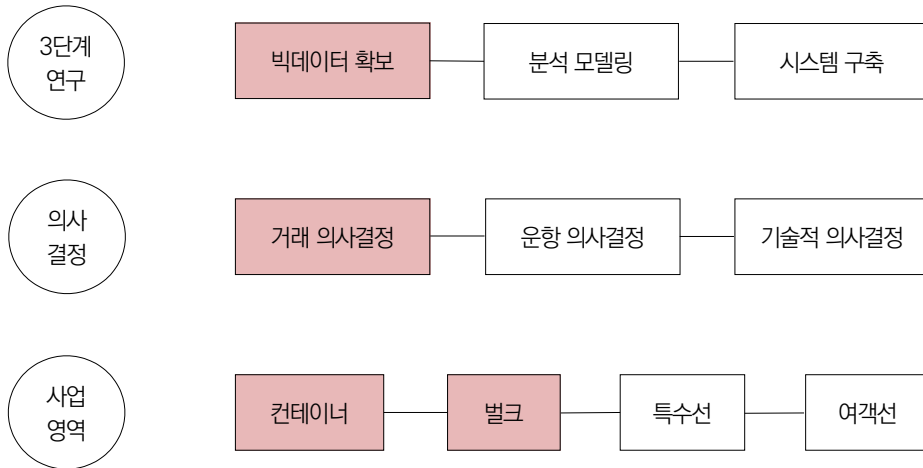
데이터의 활용에 대한 현황의 파악이 선행되어야하기 때문에 이에 대한 연구가 선행되었다. 여기서 한 가지 중요하게 언급되어야 하는 사항은 통상적으로 3V(volume, velocity, variety)로 정의되는 *빅데이터*⁸⁾의 범주에 기존의 데이터로서 다양한 원천에 산재하는 전통적 데이터를 포함하였다는 것이다. 이는 결국 시장분석을 수행하는 단계에서는 순수한 의미의 *빅데이터*가 아닌 *빅데이터*와 전통적 데이터가 모두 활용되기 때문이다.

둘째, 연구문제를 해운시장분석에 국한하였다. 해운업의 의사결정문제는 시장위험과 관련된 거래 의사결정(market decision), 운항 의사결정(operational decision), 기술적 의사결정(technical decision) 등 다양한 형태로 존재하며 이러한 문제를 포괄할 경우 데이터의 확보와 분석의 범위가 지나치게 방대해지기 때문에 성과를 낼 수 있는 단위로 나누어 문제를 해결하는 것이 중요하다. 이 연구는 시장위험과 신용위험을 다루는 시장분석으로 범위를 한정하여 수행되었다.

셋째, 사업영역의 관점에서는 컨테이너해운과 벌크해운만을 대상으로 하였으며 벌크해운과 의사결정의 유사성이 큰 선주업을 벌크해운의 범주에 포함시켜 다루었다. 따라서 액화천연가스선, 자동차운반선, 케미컬선 등의 특수선 분야와 크루즈 등 여객수송분야는 이번 연구의 범위에서 제외되었다. 이에 대한 연구는 이번 연구를 토대로 한 연구의 확장분야가 될 수 있을 것이다.

8) 이 절에서 기울임체로 표시된 빅데이터는 통상 3V로 정의되는 원래의 정의에 따른 빅데이터이다.

〈그림 1-5〉 연구의 범위



제4절 연구 방법론

해운기업의 중요한 의사결정과 데이터 활용 등의 현황을 파악하기 위해서, 국내 우수 해운기업의 경영층에 대한 인터뷰와 설문조사를 시행하였다. 설문조사는 국내 11개 해운기업을 대상으로 시행되었으며, 회수된 설문지 총 55건을 대상으로 분석을 실시하였다. 구체적인 설문조사 내용은 아래와 같다.

설문 내용은, 해운기업의 의사결정유형과 중요도에 대한 부분, 시장의 분석을 위해 활용되는 데이터와 실제 활용되는 분석기법을 파악하는 부분, 그리고 빅데이터에 대한 이해 정도를 확인하는 3부분으로 나누어 구성하였다.

설문 조사대상은, 해운기업의 영업과 기획부문에서 시장의사결정을 담당하는 다양한 계층의 조직구성원을 대상으로 하여 이루어졌다. 국내에 일정 규모의 시장분석 조직을 가진 해운기업의 수가 많지 않음을 감안하여, 일차적으로 연구진의 내부 논의 결과를 바탕으로 ‘기업’ 표본을 추출하였으며, 이차적으로 기업 내 ‘응답자’ 표본 추출은 해당 기업의 경영층에 일임하였다. 조직 외부에서는 설문에 적절히 응답 가능한 표본을 추출하기 다소 어려운 실정이었다.

또한, 빅데이터 분석에 대한 개념적인 이해와 사례 연구를 위한 광범위한 문

현 조사를 시행하였다. 아직 빅데이터 분석에 대한 명확한 개념이 정립되지 않은 것이 문헌 조사의 폭을 넓히는 요인이 되었다. 사례연구는 주로 미디어에 의존하였으며 기존의 시장분석기법이나 타 분야에서의 빅데이터 적용에 대한 연구는 주로 학술논문을 참고하였다.

데이터의 확보에 대해서는 기관의 웹사이트나 담당자 인터뷰 등이 활용되었으며 상업적인 데이터에 대해서는 데이터 공급원으로부터의 견적 또는 문의로 직접 확인하는 방법을 채택하였다.

제2장

전통적 해운시장분석과 데이터 ≪

제1절 해운과 의사결정

1. 의사결정 이론의 발전

의사결정(decision making)이란 개인이나 조직이 주어진 문제를 해결하기 위하여 여러 가지 대안(alternative)을 모색하고 선택 가능한 대안 중에서 미리 정한 기준(criteria)에 부합하는 최적의 방안을 선택·결정하는 과정을 의미한다. 여러 선행연구에서는 의사결정을 다음의 세 가지 유형으로 분류하고 있다.

김성희 외(2008)⁹⁾에 의하면 직관적 의사결정(intuitive decision making)이란 대안들과 그들의 상대적인 평가에 있어서 객관적 분석 없이 직관에 따라 행하는 의사결정이다. 일반적으로 시간이 긴박하여 다른 방법이 불가능할 때 사용되지만, 의사결정자에 따라서 대부분의 의사결정에 직관적인 판단을 선호하기도 한다. 두 번째, 기계적 의사결정(programmed decision making)은 경험적 또는 계획적 방법으로 설계되어 주어진 지침에 따라 행하는 의사결정이다. 비슷한 결정을 반복적으로 행하는 조직의 의사결정에 주로 사용되며, 필요한 노력이나 비용 면에서 효율적인 방법이다. 마지막으로 분석적 의사결정(analytical decision making)은 의사결정 이 복잡하고 대안들 간에 차이가 많을 때 사용되며, 의사결정자는 가능한 모든 대안들과 그것들의 결과를 조사하고 관련 전문가의 자문과 이용 가능한 데이터를 분석하여 가장 합리적이고 효과적인 행동경로를 선택하게 된다.

복잡한 정책결정이나 중장기 전략 계획 등 많은 요인들이 얽혀서 쉽게 접근할 수 없는 문제에 대해 인간의 인지 능력으로 결과를 예측하고 올바른 의사결정을

9) 김성희 외, 『의사결정분석 및 응용(개정판)』, 영지문화사, 2008, p.2

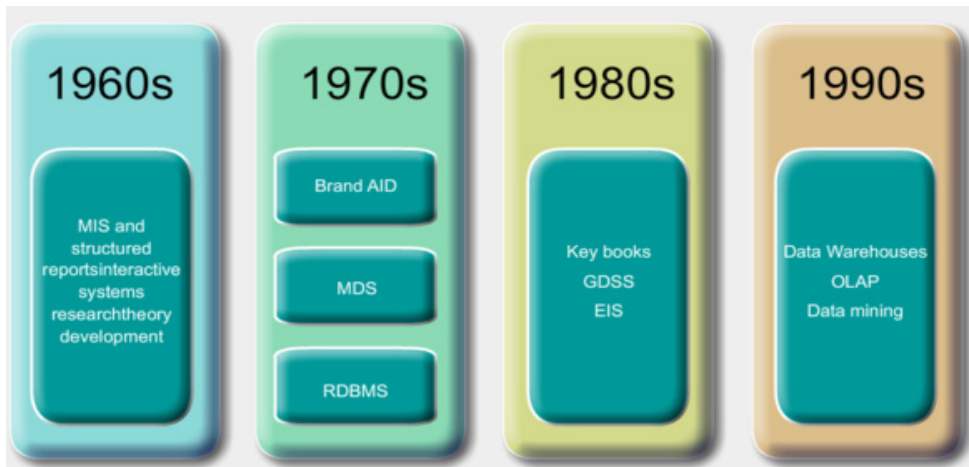
내리는 것은 불가능에 가깝다. 이와 같이 불확실성이 많이 포함된 문제를 대상으로 의사결정자의 분석적인 방법에 의한 의사 결정과정을 지원하는 의사결정 지원시스템(DSS: Decision Support System) 연구가 활발히 진행되었다. 건국대학교(2017)¹⁰⁾에 따르면 DSS라는 개념은 Peter Keen과 Charles Stabell (1978)의 연구에서 처음으로 소개된 이후 여러 학자들로부터 확대 또는 재정립되었다. 일반적으로 DSS는 경영활동에 있어 의사결정자의 비구조적(unstructured), 반구조적(semi-structured) 문제를 해결하기 위해 정보시스템의 데이터와 의사결정 모델 그리고 인터페이스를 통해 의사결정 능력을 지원해주는 컴퓨터 정보 시스템이자 커뮤니케이션 수단으로 정의할 수 있다. Keen과 Stabell은 1950년대 후반에서 60년대 초 카네기공과대학에서 수행된 조직의사결정에 대한 이론적 연구와 1960년대 매사추세츠공과대학에서 실행된 상호작용적인 컴퓨터 시스템에 관한 기술적 연구를 중심으로 DSS의 개념을 정립하였다. 1979년 John Rockart는 하버드 비즈니스 리뷰에 중역정보시스템(EIS: Executive Information System) 또는 중역지원시스템(ESS: Executive Support System)의 개발을 가져온 혁신적 기사를 게재하였으며, 1980년 Steven Alter는 저서에서 사례연구를 통해 DSS의 개념을 소개하였다. Bonczek(1981)은 DSS 설계와 관련된 문제를 이해하기 위한 이론적 체계를 정립하였으며, Sprague와 Carlson(1982)은 조직이 DSS를 구축 할 수 있는 방법과 실천해야 할 행동에 대해 실용적이고 이해하기 쉽게 설명하고 있다. 1980년대 중반에는 많은 학계 연구자들이 조직의 의사결정을 지원하는 소프트웨어인 GDSS(Group Decision Support System)를 개발하였으며, GDSS의 영향과 결과에 대한 연구 조사가 뒤따랐다. 그 중 EIS는 단일사용자 대상 모델기반의 의사결정지원시스템으로 향상된 관계형 데이터베이스 제품으로 발전하였다.

1990년대에는 데이터 웨어하우징(data warehousing)과 온라인 분석프로세싱(OLAP: On-Line Analytical Processing)이 DSS의 영역을 확장하였으며, 데이터 기반 DSS의 범위를 더욱 넓혔다. 데이터 웨어하우징이란 개방형 시스템

10) 건국대학교 경영대학원, 의사결정지원시스템의 개요, <https://home.konkuk.ac.kr:8080/cms/Common/MessageBoard/ArticleFile.do?id=53c9c32>, 검색일 : 2017.12.05

도입으로 흩어져 있는 방대한 양의 데이터에 최종 사용자가 쉽게 접근하고 활용할 수 있는 신속한 의사결정에 활용할 수 있게 하는 기술로 정의된다.¹¹⁾ 반면 OLAP은 다차원 데이터 구조를 이용하여 다차원의 복잡한 질의를 고속으로 처리하는 데이터 분석 기술이다.¹²⁾ 전산 처리 및 저장 기술의 급속한 발전으로 90년대 후반에는 데이터마이닝(data mining) 기술이 본격적으로 실제 기업의 의사결정에 활용되기 시작하였다.

〈그림 2-1〉 의사결정지원시스템(DSS)의 발전단계



자료 : Richard Kása, Knowledge management, 2014

DSS의 개발과 연구는 타 학문 분야와 연계하여 발전되어 왔다. 데이터베이스 전문가들은 대용량 데이터와 문서 처리 기술과 방법론을 제공하였으며, 경영과학은 DSS에서 사용되는 수학적 모델을 개발하고 문제해결을 위한 모델링의 장점을 증명하였다. 인지과학, 특히 행동 의사결정 연구는 DSS 설계에 도움이 되는 설명적인 정보를 제공해 왔고 DSS 연구를 위한 가설을 생성해 왔다. 다른 중요한 관련 분야로는 인공지능(artificial intelligence), 인간과 컴퓨터 간 상

11) 매일경제용어사전, <http://terms.naver.com/entry.nhn?docId=3360&cid=43659&categoryId=43659>, 검색일 : 2017.12.05

12) 두산백과, <http://terms.naver.com/entry.nhn?docId=2807097&cid=40942&categoryId=32840>, 검색일 : 2017.12.05

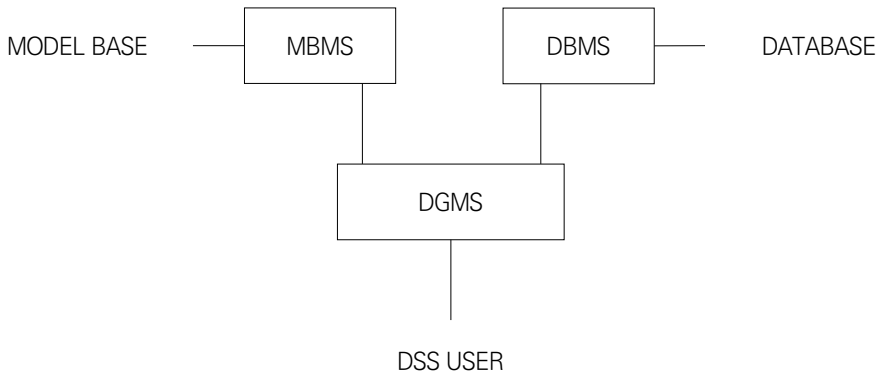
호작용(human-computer interaction), 시뮬레이션 방법(simulation methods), 소프트웨어 공학(software engineering), 그리고 텔레커뮤니케이션(telecommunications) 등이 있다. 최근에는 계량적/통계적/수학적 모델링 기법으로 특정 의사결정 문제를 상세히 분석하는 어널리틱스(analytics, analytical DSS)가 일반화되어 확산되고 있다.¹³⁾

DSS는 일반적으로 데이터베이스(database) 시스템, 모델베이스(model-based) 시스템, 사용자 인터페이스(user interface) 등의 요소로 구성된다. 데이터베이스 시스템은 다양한 데이터를 저장하고 있는 데이터베이스와 이를 관리하는 데이터베이스 관리시스템(DBMS: Database Management System)으로 구성되며, 의사결정에 필요한 데이터를 저장·관리하고 이를 제공하는 기능을 한다. 모델베이스 시스템은 다양한 모델을 저장하고 있는 모델베이스와 이를 관리하는 모델베이스 관리시스템(MBMS: Model-based Management System)으로 구성되며, 의사결정에 필요한 모델을 개발·수정·통제하는 기능을 제공한다. 마지막으로 사용자 인터페이스는 데이터의 입·출력과 다양한 분석과정에서 일어나는 사용자와 시스템 간의 인터페이스 환경을 제공하는 시스템모듈로서 주로 메뉴방식이나 그래픽 처리형식을 이용하여 사용자가 쉽게 이해하고 사용할 수 있는 대화기능을 제공하기 때문에 대화생성 및 관리소프트웨어(DGMS: Dialogue Generation and Management Software)라고도 한다. 이밖에 DSS의 구성요소로 정성적인(qualitative) 정보를 제공하는 지식베이스 시스템을 포함시키기도 한다.¹⁴⁾

¹³⁾ 건국대학교 경영대학원, 의사결정지원시스템의 개요, <https://home.konkuk.ac.kr:8080/cms/Common/MessageBoard/ArticleFile.do?id=53c9c32>, 검색일 : 2017.12.05

¹⁴⁾ 위키백과(https://ko.wikipedia.org/wiki/의사결정_지원_시스템) 검색 내용을 바탕으로 저자 정리

〈그림 2-2〉 의사결정지원시스템(DSS)의 구조



자료 : Andrew P. Sage, Decision Support Systems Engineering, 1991

DSS는 의사결정과정을 비용 중심의 효율적(efficiency)인 면보다 목표 중심의 효과적(effective)인 측면에서 향상시킨다고 할 수 있다. DSS를 사용하는 가장 주된 이점은 보다 좀 더 나은 의사 결정을 하는 것이다. 즉, 문제 분석을 통해 다양한 대안들을 제시해 정해진 기준 하에 최적의 대안을 선택하는 과정을 효과적으로 지원하는 것이다.¹⁵⁾ 이를 통해 장기적인 수익성을 향상시키고, 외부의 환경 변화에 신속하게 대응할 수 있게 한다. 또한 내부의 강·약점, 외부 위협 및 기회에 대한 의사결정자의 인식을 개선하고, 의사결정 경험을 축적하여 미래에 발생할 수 있는 유사한 상황에 사용할 수 있게 한다.

2. 해운기업의 의사결정

해운업의 중요 전략적 의사결정 문제들은 비구조적(unstructured)이나 반구조적(semi-structured) 문제의 형태를 띠고 있으며 대부분 분석적 의사결정(analytical decision making)을 요하는 문제들이다. 이러한 의사결정의 경우 합리성과 객관성을 확보하고 의사결정의 질을 높이기 위해서는 데이터에 기반을 둔 의사결정 지원시스템의 확보가 필수적이다. 전산능력의 비약적 발전으로

¹⁵⁾ 건국대 경영대학원, <https://home.konkuk.ac.kr/> 의사결정시스템의 개발

최근에 그 가치가 재조명되고 있고 적용의 범위를 넓히고 있는 인공지능 기반의 도구들이 활용되는 소위 어널리틱스(analytics)의 적용이 필요한 영역이 해운이다.

이어서 해운사업 영역별로 어떤 의사결정이 이루어지는지 파악해 보기로 한다. 해운업에는 다양한 사업모델이 있고 사업모델 별로 소위 ‘게임의 규칙’이 크게 다르기 때문에 의사결정을 동일한 관점에서 다루기 어렵다. 물론 수송수단을 매개로 해서 이윤을 창출한다는 공통점이 있기 때문에 선사로서 공통적으로 의사결정 하는 영역이 있지만 사업모델별로 구분해야 할 필요가 있다.

해운의 사업모델은 분류의 기준에 따라 다양하게 나누어질 수 있으나 가장 일반적으로 분류하는 것은 선주업과 운송업을 나누고 다시 운송업을 컨테이너와 벌크선으로 나누는 것이다. 아래는 컨테이너와 벌크의 차이점을 표로 정리한 것이다.

〈표 2-1〉 벌크해운과 컨테이너해운의 비교

사업속성	컨테이너해운	벌크해운
거래	• 선박투자 또는 장기용선의 형태로 수송 능력 ¹⁶⁾ 을 일괄 구매한 후 TEU단위로 분할 매각	• 선박투자 또는 다양한 용선계약을 통해 수송 능력을 확보한 후 다양한 계약형태로 매각
경쟁우위요소	• 규모의 경제 • 운영효율성 (연료효율, 하역효율 등)	• 거래시점, 형태, 기간의 선택 • 시장위험관리
역량	• 조직역량 • 금융 역량 • IT 역량	• 인적역량 • 의사결정역량
시장위험 (Market risk)	• 가격수용자(price taker)이며 단일시점 일괄구매의 영향으로 시장위험에 완전 노출	• 거래 시점과 기간에 따른 가격의 가변성과 거래형태의 다양성으로 노출 조절 가능

자료 : 저자정리

여기서는 컨테이너와 벌크를 나누고, 의사결정의 유사성을 감안하여 선주업을 벌크선과 같은 속성으로 놓고 의사결정을 정의하였다.

¹⁶⁾ 수송능력(cargo-carrying capacity)은 선박의 수송공간에 해당 공간의 사용시간을 곱한 것으로 정의한다.

1) 공통 의사결정

해운업의 유형에 관계없이 내리는 의사결정으로서 가장 중요한 것이 사업 포트폴리오에 대한 의사결정이다. 이는 전략적 사업단위(SBU: Strategic Business Unit)라고도 하는 단위사업에 진입할 것인가, 철수할 것인가, 단위사업을 확장할 것인가, 축소할 것인가의 네 가지 하부 의사결정으로 구성된다. 이러한 의사결정들은 철저하고 심도 있는 시장분석과 내부 역량분석에 기초하여 이루어져야 한다.

사업을 영위하는 데에 또 하나의 중요한 의사결정은 신용위험과 관련된 것이다. 거래상대방의 건전성을 평가하는 것은 해운과 같이 거래상대방의 법적 정정성이 분명하지 않은 산업에 있어서는 무엇보다도 중요하다. 한국의 해운기업들도 불황기에 거래 상대방의 신용이 붕괴됨으로써 많은 어려움에 직면하였고 선박금융기관 또한 거래 상대방의 신용상태가 변화된 것을 감지하지 못해 불황기에 많은 부실자산을 떠안게 되었다.

2) 컨테이너 사업의 의사결정

컨테이너사업의 고유의 의사결정은 크게 항로기획과 세일즈로 나누어 볼 수 있다. 항로기획은 컨테이너에 있어서의 항로단위 포트폴리오 전략이라고도 할 수 있다. 즉, 항로별로 진입, 철수, 확대, 축소를 결정하는 것이다. 이 의사결정은 투입선형과 서비스 빈도의 결정이 핵심이 되며 한번 의사결정이 내려지면 그 영향이 장기에 미치기 때문에 다양한 영향요인이 충분히 고려되어야 한다. 그 중에서도 시장분석 즉, 수요, 공급(경쟁), 가격에 대한 분석과 예측은 항로기획 성패의 핵심요소가 된다.

세일즈에 대한 의사결정은 주로 가격의 결정과 고객군의 구분 및 선정과 관련된 것이기 때문에 노선의 기반구조를 결정하는 항로기획보다는 단기적이고 가변적인 의사결정이 된다. 특히 컨테이너에 있어 선사는 가격수용자 (price taker)의 입장이기 때문에 가격결정에 대한 재량권은 매우 제한적이라고 할 수 있다.

3) 벌크선 사업과 선주업의 의사결정

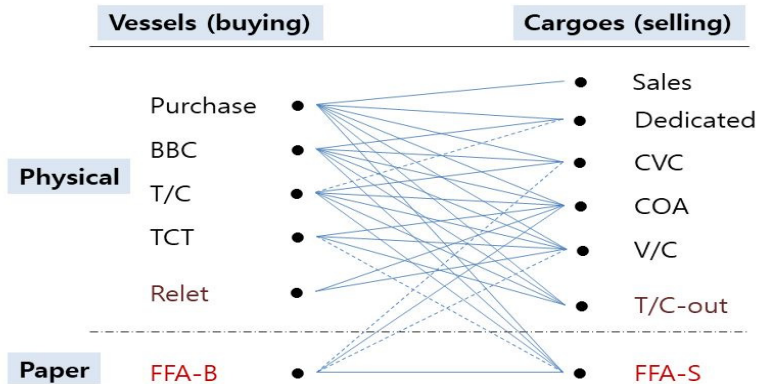
컨테이너 사업이 수송수단을 확보하고 시간의 축을 따라가며 분할하여 슬롯(slot)을 매각하는 획일적인 구조를 가진 반면 벌크선 사업은 훨씬 동적(dynamic) 속성을 가지고 있다. 이는 그림 2-3의 트레이딩 구조에서 보듯이 수송능력의 구매와 판매의 형태가 다양하고 거래의 규모가 되는 기간도 다양하게 설정되기 때문이다.

벌크선 사업 의사결정의 핵심은 수익과 원가에 변화를 가져오는 거래시점과 거래기간을 결정하는 것이다. 특히 거래기간을 결정하는 것은 벌크해운에 있어서 매우 중요한 부분이다. 운임시장의 특성 상 단기물의 변동성이 장기물보다 크기 때문에 거래기간의 선택이 가격의 결정과 직결되기 때문이다.

결국 의사결정의 본질은 저가에 구매하고 고가에 판매한다는(buying low, selling high) 트레이딩의 기본을 통해 이윤을 극대화하는 것이다. 이 과정에서 필수적으로 수반되는 의사결정 영역이 시장위험에 대한 의사결정이다. 벌크선사업의 경우 시황변동에 대한 노출의 조정이 비교적 자유롭기 때문에 순노출의 정도가 기업이 감당할 수 있는 범위를 초과하게 되면 기업은 도산의 위험에 내몰리게 된다. 실제 시황 급등기에 단기적인 이익을 추구하여 노출을 증대시킨 많은 기업들이 뒤이어 도래하는 하락기에 도산하거나 위기를 맞았다.

선주업의 경우에도 아래 그림의 선박구매와 대선, 선박매각의 조합이기 때문에 일반 벌크선 사업의 트레이딩과 유사한 구조를 가지고 있다. 따라서 의사결정의 내용이 같다고 보아도 무방할 것이다.

〈그림 2-3〉 해운업의 트레이딩 구조



출처 : Yun, H., et al. (2017).

4) 해운기업 의사결정의 현황

해운기업의 주요 의사결정을 다음의 5가지로 분류하고, 설문조사를 통하여 중요도의 우선순위를 파악했다.

- 신규사업진입, 사업확장, 사업축소, 사업철수, 전략적 제휴 등 사업구조의 결정
- 선박의 확보(선박구매, 용선 등) 및 선박의 판매(수송계약, 대선 등)의 시점 및 계약기간의 선택
- 거래처의 선정 및 신용의 판단
- 선속의 증가, 선원비 절감 등 운영효율(투입 대비 산출)제고 방안의 선택
- GRI 등 판매 및 구매가격의 결정

설문분석을 실행한 결과, 전체 응답자의 65%에 해당하는 36명이 첫 번째 항목인 ‘사업구조의 결정’을 의사결정 1순위로 선택하였다. 이는 해운기업에 있어서 신규사업진입, 사업확장, 사업축소, 사업철수, 전략적 제휴 등 사업구조의 결정을 가장 중요한 의사결정으로 인식하고 있음을 알 수 있다. 다음으로 선박매매시점 및 계약기간의 선택을 1순위로 한 응답자가 총 15명으로, 전체의 27%에 달했다. 그 외, 신용위험의 문제, 운영효율의 문제, 구매가격 결정 등을 가장 중요한 의사결정으로 선택한 응답자는 총 4명으로 전체의 7%에 불과했다. 전체

응답자의 92%가 1순위로 선택한 사업구조의 결정 및 선복매매시점, 계약기간의 선택문제 등은, 향후 빅데이터 해운시장분석에 있어서, 해운기업의 의사결정문제를 정의할 때 우선적으로 다루어져야 할 것이다. 표2-2는 중요도에 따른 해운기업 의사결정 우선순위에 대한 설문조사 결과를 표로 정리한 것이다.

〈표 2-2〉 중요도에 따른 해운의사결정의 우선순위

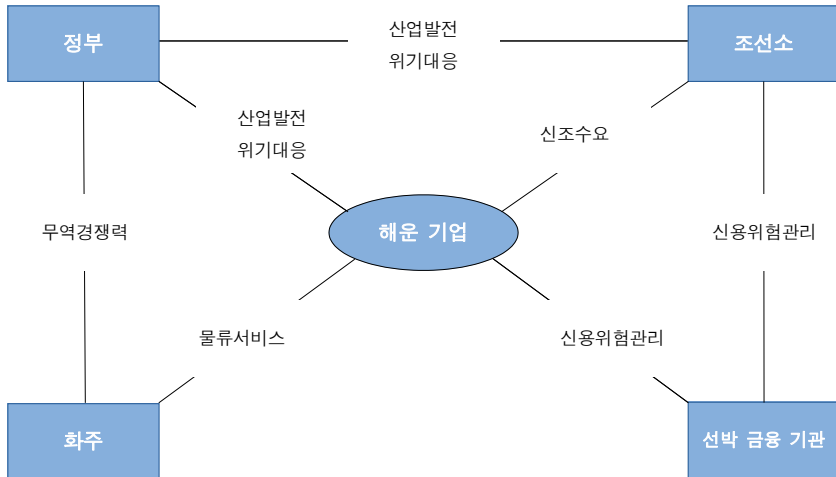
단위: 명

	1순위	2순위	3순위	4순위	5순위
사업구조의 결정	36	10	3	3	1
선복의 확보 및 선복의 판매의 시점 및 계약기간의 선택	15	30	7	2	0
거래처 선정 및 신용판단	3	5	21	15	8
선속의 증가, 선원비 절감 등 운영효율 제고방안의 선택	1	5	9	23	14
GRI 등 판매 및 구매가격의 결정	0	3	14	7	22

3. 해운의사결정의 산업간 연관구조

해운업의 의사결정은 관련된 여러 분야에 영향을 미친다. 그림 2-4는 관련 산업 또는 기관 간의 연관구조를 나타낸 것이다.

〈그림 2-4〉 해운 의사결정의 연관구조



자료: 저자 정리

조선소는 해운의 수요를 기반으로 하는 산업이다. 통상적으로 호황기에 선박의 발주가 증가하고 불황기에는 발주가 감소하여 조선 산업이 위기에 처하게 되므로 해운업과 조선업은 동일한 의사결정 기반을 갖는다고 할 수 있다. 특히 시황의 예측이나 경기변동 국면의 판단에 있어서는 분석에 사용되는 데이터나 분석기법에 거의 차이가 없을 것으로 판단된다. 조선소는 이러한 예측과 판단에 기초하여 수주 가격을 결정하거나 기간별 건조도크 선매도 여부를 결정하는 등 핵심적인 전략적 의사결정을 하게 된다.

금융기관은 기업에 타인자본 형태의 자금을 제공하는 주체로서 주주와 더불어 기업의 양대 이해관계자가 된다. 대출의 목적이 선박금융이든 운전자금이든, 신용의 공여가 이루어지기 이전이든 이루어진 이후이든 선박금융기관은 차주의 상환불능에 따르는 손실을 최소화하기 위하여 신용위험관리에 초점을 맞추게 된다. 신용위험관리는 대부분 재무제표에 대한 분석을 기초로 이루어지지만 이는 사후적이며 정태적인 성격을 갖는 한계가 있다. 사전적이고 동태적인 신용위험관리를 위해서는 해운기업 시장위험관리의 틀이 적용될 필요가 있다. 즉, 해운업 의사결정을 위한 데이터분석을 활용할 경우 위험에 선제적으로 대응하는 것이 가능해진다. 선박금융기관은 시장분석의 결과를 사전적으로 대출여부를 결

정하거나 이자율을 결정하는 데 활용할 수 있으며, 사후적으로는 기업의 부실에 대한 예측에도 활용할 수 있다. 특히 선박금융기관이 전통적으로 호황기에 선사가 고가로 수송수단을 확보하는 것을 조장한다는 비난에 직면해 있어 분석의 결과를 시장 과열 시 대출 억제에 근거로 활용한다면 해운시장의 변동성 축소와 신용위험의 감소라는 양면의 효과를 기대할 수 있을 것이다.

화주는 선사로부터 해상운송서비스를 제공받는 주체이므로 해운의 수익은 화주의 입장에서는 물류비용이 된다. 기업에 있어 물류비용이 전체 제품원가의 약 10%를 차지하고 있고 그 중 해상운송비용의 비중이 크므로 해운시장분석의 중요한 이용자이다. 화주가 운임시황의 변동에 대응해서 적극적으로 물류비용을 통제할 것인지 아닌지는 화주의 위험회피성향에 달려있지만 대부분의 화주들은 시황을 따라가며 비용을 지불하기 보다는 수송시점과 계약기간을 합리적으로 선택함으로써 물류비용을 절감하고자 한다. 컨테이너화주의 경우 비교적 고가의 화물을 수송하므로 시점과 기간 선택의 여지가 작으나 컨테이너화주라고 하더라도 대량화주의 경우에는 기간의 선택 즉 장기수송계약을 통해 일정기간 운임을 고정함으로써 수송비를 절감할 수 있다. 이와는 달리 벌크화주의 경우에는 시점과 기간을 비교적 자유롭게 선택하여 수송비용에 큰 변화를 줄 수 있다. 컨테이너 운임시황에 비하여 벌크시황의 단기변동성이 큰 것도 시점선택의 효과를 크게 하는 요인 중의 하나이다.

해운의 의사결정과 그에 따른 경영성과에 큰 영향을 받는 또 하나의 주체는 정부이다. 정부는 시장의 과열에 따르는 자원배분의 비효율을 예방하고 불황기에 대규모의 기업실패가 일어나지 않도록 지원하는 산업정책의 주체이다. 따라서 해운기업 시장분석의 틀을 공유하여 끊임없이 분석하고 예측해야 하며 특히 조기경보시스템 등 문제의 징후를 사전에 인지할 수 있는 시스템을 운영함으로써 사전적 대응능력을 갖추고 있어야 한다.

제2절 해운시장분석과 전통적 데이터

해운시장분석은 크게 정량적 시장분석과 정성적 시장분석으로 나눌 수 있다. 정량적 시장분석의 영역에서는 운임, 선가 등 가격변수의 동태적 특징에 대한 연구가 많이 이루어졌다. 본 연구에서는 정량적 시장분석을 전통적 시계열 분석과 복잡계 모형으로 나누어 지금까지의 연구를 소개하고 정성적 분석이 이루어지는 절차를 살펴본다. 이어서 이러한 시장분석에 활용되는 전통적 데이터의 종류와 활용 현황에 대해 설문조사 결과를 분석한다.

1. 정량적 시장분석

1) 전통적 시계열 분석

우선 정량적으로 해운시장을 분석한 선행연구로서 Beenstock and Vergottis (1993)¹⁷⁾의 연구를 언급할 필요가 있다. 이들은 건화물선, 유조선, 건화물/유조선 겸용선에 대해 운송시장의 수요(물동량), 공급(선박량), 가격(운임) 등을 고려하고, 신조선 시장, 중고선 시장, 해체 시장 등도 포함한 일반균형이론을 제시하였으며 계량경제모형으로 실제 자료를 가지고 모수(parameter)를 추정했다. 또한 제시된 모형과 추정된 모수 값을 기초로 수요, 공급 등의 다양한 부문에서의 충격이 해운시장에 미치는 영향을 시뮬레이션 기법을 통해 분석하고 있다.

이와 같은 복잡한 일반균형 관점에서의 모형개발에도 불구하고, 역설적이게도 Beenstock and Vergottis(1993)의 연구 이후 해운시장의 구조적 특징을 반영한 연구는 학계의 주요 논의대상이 되지 못하였다. 후술하겠지만 학계에서는 오히려 축약된 형태의 계량경제모형, 예를 들어 공적분(cointegration) 개념의 VECM 등이 광범위하게 연구되기 시작했다.

Kavussanos and Nomikos (2003)¹⁸⁾는 파나마스 스팟 운임과 BIFFEX 선

¹⁷⁾ Beenstock, M. and A. Vergottis (1993), *Econometric Modelling of World Shipping*, Chapman & Hall.의 자료를 저자가 정리하여 작성하였다.

¹⁸⁾ Kavussanos, M. G. and N. K. Nomikos (2003), "Price Discovery, Causality and Forecasting in the Freight Futures Market", *Review of Derivatives Research*, 6, pp. 203-230.의 자료를 저자가 정리하여

물 운임 간의 관계 분석 및 그에 기초한 운임 예측에 벡터오차수정모형(VECM: Vector Error Correction Model)을 적용했다. 그들은 해운서비스 시장에서 전통적인 상품(commodity)과는 달리 서비스라는 저장 불가능한 상품이 거래되기 때문에 선물 가격이 소위 보유비용(cost of carry)을 반영하여 현물 가격과 연동되지 않는 근본적 차이가 있다는 점에 주목했다. 즉, 해운서비스 시장의 선물 운임(BIFFEX 선물 운임)이 현물 운임에 대한 가격예시기능(price discovery)을 하지 못할 수 있다는 것이다. 또한 벡터오차수정모형과 함께 이들은 ARIMA, 벡터자기상관모형(VAR), 임의보행모형(random walk model)의 예측력을 비교 분석했다. 이들의 분석에 따르면 파나마스 스팟 운임과 선물 운임은 장기적인 선형 관계를 가지는 것으로 나타나, 벡터오차수정모형을 이들 시장의 분석에 사용할 수 있음을 보이고 있다. 또한 벡터오차수정모형의 그랜저 인과성 검증(Granger causality test)과 충격반응분석(impulse response analysis)을 통해 선물 가격을 현물 가격의 예측에 사용하는 것이 유용함을 보이고 있다. 즉 선물 운임이 현물 운임의 가격예시기능을 한다는 것이다.

Kavussanos, Visvikis and Menachof (2004)¹⁹⁾는 파나마스 월간 스팟 운임과 FFA 가격 자료를 활용하여 소위 선물 가격의 현물 가격에 대한 “불편추정량 가설”(unbiasedness hypothesis)을 검증했다. Kavussanos and Nomikos (2003)의 문제 의식과 같이 이들은 해운서비스는 저장이 불가능하기 때문에 선물가격에 보유비용을 반영하여 선물운임이 형성되지 않다는 점에서 가설 검증의 필요성을 제기했다. 나아가 이들은 해운서비스의 스팟 시장과 선물 시장은 거래 비용에 있어 비대칭적이라는 사실에 주목했다. 즉, 스팟 시장에서 해운서비스를 제공하기 위해서는 선박을 선행적으로 확보해야 하기 때문에, 이러한 선박 확보 비용이 소요되지 않는 선물 시장에서 거래비용이 현물 시장의 거래 비용보다 유의미하게 낮다는 것이다. 따라서 두 시장이 해운시장에 가해지는 충격에 반응하는 패턴이 다르다고 추정할 수 있다.

작성하였다.

¹⁹⁾ Kavussanos, M. G., I. D. Visvikis and D. Menachof (2004), “The Unbiasedness Hypothesis in the Freight Forward Market: Evidence from Cointegration Tests”, *Review of Derivatives Research*, 7, pp. 241-266.의 자료.

이들 분석결과에 따르면, 대서양 수역의 3개월 만기물을 제외한 모든 시장에서 선물 운임은 스팟 운임의 불편 추정량이 된다는 귀무 가설을 기각할 수 없는 것으로 나타났다. 또한 대서양 수역의 3개월 만기물에서 불편 추정량 가설이 기각된 사실에 근거하여 선물의 만기가 길수록 스팟 시장과 현물 시장의 움직임이 분리될 가능성이 크다고 해석하고 있다. 또한 선물 운임이 과거의 불균형에서 균형으로 돌아가는 경향이 크다는 점에서 선물 시장의 낮은 거래비용이 이 같은 결과를 나타낸 원인으로 해석될 수 있다. 즉, 해운시장에 충격이 왔을 때 선물 시장은 상대적으로 거래비용이 적기 때문에 이러한 충격에 보다 쉽게 반응한다는 것이다.

Batchelor, Alizadeh and Visvikis (2007)²⁰⁾는 ARIMA, VAR, VECM, 제약된 VECM 모형을 일간 파나마스 운임 자료에 적용하고 있다. 앞서의 두 연구와 마찬가지로 이들도 해운시장에서 스팟 시장과 선물 시장 간의 장기적 관계를 다루고 있다. 이들의 분석 결과에 따르면 다음과 같은 시사점이 도출된다. 첫째, 선물 운임이 스팟 운임에 비해 모형을 통한 예측이 어렵다. 그러나 장기 균형에 수렴하는 속도 측면에서 보면, 선물운임의 균형 회복 속도가 현물운임보다 빠르다. 둘째, VECM 모형은 임의보행모형에 비해 예측력이 뛰어나다. 이 같은 예측력 개선은 단기 예측 보다는 장기 예측에 있어 더욱 두드러진다. 셋째, 스팟 운임 예측에 있어 VECM 또는 제약된 VECM 모형이 VAR 모형보다 예측력이 뛰어나다. 특히 이 같은 VECM 모형의 우수성은 선물 시장의 유동성이 대서양 항로에 비해 더 큰 태평양 항로의 시장에서 더욱 유의미하게 나타났다.

Alizadeh, Adland and Koekebakker (2007)²¹⁾는 이전의 연구들이 선물 시장의 자료를 활용한데 비해, 용선시장의 자료를 사용했다는 점에서 차이가 크다. 벌크선 시장에서는 운송시장의 스팟 운임의 기대치 등이 반영된 기간용선시장이 활발히 활용되고 있다. 따라서 효율적 시장 가설(efficient market hypothesis)에 따르면 1년 용선료에서 6개월 용선료를 제한 나머지 용선료는

20) Batchelor, R., A. Alizadeh and I. Visvikis (2007), "Forecasting Spot and Forward Prices in the International Freight Market", *International Journal of Forecasting*, 23, pp. 101-114.

21) Alizadeh, A. H., R. O. Adland and S. Koekebakker (2007), "Predictive Power and Unbiasedness of Implied Forward Charter Rates", *Journal of Forecasting*, 26, pp. 385-403.

미래의 6개월 용선료가 되어야 한다. 이들은 이렇게 계산된 미래 용선료를 암묵적 미래 용선료(Implied Forward Time Charter)라 부르고 이들 IFTC가 실제 미래 용선료를 얼마만큼 정확히 예측하는지를 분석했다.

이들의 분석결과에 따르면, IFTC 예측은 ARIMA, VAR, VECM 기법 보다 개선된 예측력을 보여 주었다. 그러나 용선전략으로서 이 같은 IFTC에 기초한 전략은 단순 이동평균(moving average) 전략에 비해 열위에 있다는 다소 모순된 결과를 도출했다. 또한 이들은 시장이 상승기에 IFTC가 미래 용선료를 과소 예측하는 경향이 있다는 사실을 언급하고 있다. 즉, 시장의 상태에 따라 IFTC의 예측에 체계적인 오류가 있다는 사실을 알 수가 있다.

2) 복잡계 모형

Dikos, Marcus, Papadatos and Papakonstantinou (2006)²²⁾는 유조선 운송시장의 모형화에 시스템 다이내믹스를 활용했다. 시스템 다이내믹스 기법은 해운시장의 다양한 상황의 특징을 반영하면서, 경영 의사결정과 여러 가지 시나리오의 평가에 대한 툴(tool)로서 기능한다고 평가된다. Randers and Goluke (2007)²³⁾ 또한 유조선 시장에 시스템 다이내믹스 기법을 적용했다. 그들의 연구 결과에 따르면, 시스템 다이내믹스 모형은 해운시장에서의 변곡점(turning point)을 예측하는데 성공한 것으로 보고되고 있다.

Li and Parsons (1997)²⁴⁾은 인공 신경망 모형(artificial neural networks)을 유조선 시장 분석에 적용했다. 이들은 유조선 시장의 월간 스팟 운임, 물동량, 선박량 등으로 구성된 자료를 활용했다. Lyridis, Zacharioudakis, Mitrou and Mylonas (2004)²⁵⁾ 또한 인공 신경망 모형을 유조선 시장에 적용한 바 있

22) Dikos G., H. Marcus, M. P. Papadatos and V. Papakonstantinou (2006), "Nivers Lines: A System-Dynamics Approach to Tanker Freight Modeling", *Interface*, Vol. 36, No. 4, pp. 326-341.

23) Randers, J. and U. Goluke (2007), "Forecasting Turning Points in Shipping Freight Rates: Lessons from 30 Years of Practical Effort", *System Dynamics Review*, Vol. 23, No. 2/3, pp. 253-284.

24) Li, J. and M. G. Parsons (1997), "Forecasting Tanker Freight Rate Using Neural Networks", *Maritime Policy and Management*, Vol. 24, No. 1, pp. 9-30.

25) Lyridis, D., P. Zacharioudakis, P. Mitrou and A. Mylonas (2004), "Forecasting Tanker Market Using Artificial Neural Networks", *Maritime Economics and Logistics*, pp. 93-108.

다. 이들 인공 신경망 모형은 복잡한 시계열 자료 간의 불안정한 관계를 잘 반영하는 것으로 평가된다.

2. 정성적 시장분석

해운시장분석에서 정성적 분석은 해운기업과 정부, 관련 기관 등의 의사결정에 매우 중요한 역할을 한다. 특히 해운산업은 자본집약적 서비스 산업으로서 기업들의 전략적 의사결정이 시장의 행태와 성과에 미치는 영향이 매우 크다. 여기서는 Martin Stopford(2009)²⁶⁾가 소개하는 시장조사 방법론을 간략히 살펴본다.

시장조사 방법론은 다음과 같은 6단계로 구성된다: i) 위임사항의 결정, ii) 과거추세의 분석, iii) 경쟁자의 계획과 전문가의 의견 조사, iv) 미래 시장 발전에 대한 영향의 구분, v) 정보를 예측으로 결합, vi) 결과의 제출.

첫째, 구체적인 시장조사에 앞서 이 같은 조사를 수행하는 의미를 분명히 해야 한다. 예를 들어, 어떤 의사결정이 이루어질 것이며, 이 조사가 어떠한 기여를 하게 될 것인지를 명확히 정의해야 한다. 둘째, 과거 추세를 분석하는 작업에서는 조사의 대상이 되는 시장을 세분화하여 정의해야 한다. 이 같은 시장 정의가 선행되면, 다음으로 이용 가능한 모든 정보를 모으고 과거 추세를 분석한다. 셋째, 시장이 정의되고 과거 추세가 분석되면, 다음으로 경쟁자를 구분하고 경쟁자의 계획과 전략을 파악해야 한다. 특히 이 같은 경쟁자의 전략 분석의 모범 사례로 한국기업평가의 서강민·김종훈(2016)²⁷⁾ 보고서는 매우 좋은 사례를 제시한다. 이들은 머스크가 과점적인 글로벌 컨테이너선 시장에서 가격선도자로서 역할을 하면서 운임수준 등의 결정에 매우 중요한 역할을 하고 있음을 다양한 자료를 통해 설명하고 있다. 이 같은 전략 분석은 정량적 분석의 한계를 보완하여 미래 시장 움직임에 대한 보다 정확한 정보를 제공하는 것으로 평가된다.

²⁶⁾ Stopford, M. (2015), 『해운경제학』 (양창호 외 3인 역).

²⁷⁾ 서강민·김종훈 (2016), “해운, 2차 치킨게임의 서막 - 현대상선, 용선료 인하 이후 생존가능성은?”, 『한국기업평가』.

한편 위의 1~3단계의 시장 조사는 광범위한 데이터베이스의 구축을 요구한다. 즉 데이터베이스의 구축을 통해 시장에서 일어나고 있는 사건들의 전체 모습을 제공하고 분석전문가가 조사하고 설명할 수 있는 기초자료를 제공할 수 있게 된다. 한편, 전문가의 의견조사는 면접조사, 설문조사 등의 방법을 통해 수행될 수 있다.

넷째, 미래 시장 발전에 대한 세부 시장 간의 영향 등을 분석해야 한다. 해운 시장에서는 크게 화물과 선박의 종류 및 크기에 의해 분할하여 세분화된 사업을 영위할 수 있지만, 이 같은 분할이 전적으로 세부 시장의 배타적 구분을 의미하는 것은 아니다. 예를 들면, 철광석 물동량이 증가하면 이는 벌크선의 케이프, 파나막스, 수프라막스에 그 크기는 다르지만 영향을 미치게 된다. 따라서 이러한 미래 변수들이 시장에 어떠한 영향을 미칠지에 대한 검토가 필요하다.

다섯째, 이상의 4단계까지의 시장 조사가 일단락되면 수집되고 분석된 정보를 활용하여 미래 상황을 예측해야 한다. 물론 이 같은 미래 상황 예측은 시장 조사에 위임된 내용을 충족하는 것이어야 한다. 마지막으로 이와 같은 시장조사의 내용은 보고를 받는 사람의 필요에 맞게 가공되어 보고되어야 할 것이다.

3. 전통적 데이터의 유형

설문조사를 통하여 해운기업에서 실제 시장분석에 활용하는 데이터의 공급원천을 조사하였다. 내·외부 전문가를 사전 인터뷰한 내용을 토대로 아래의 유형과 항목을 제시하였으며 제시되지 않은 경우 유형과 항목을 추가로 기술하도록 하였다.

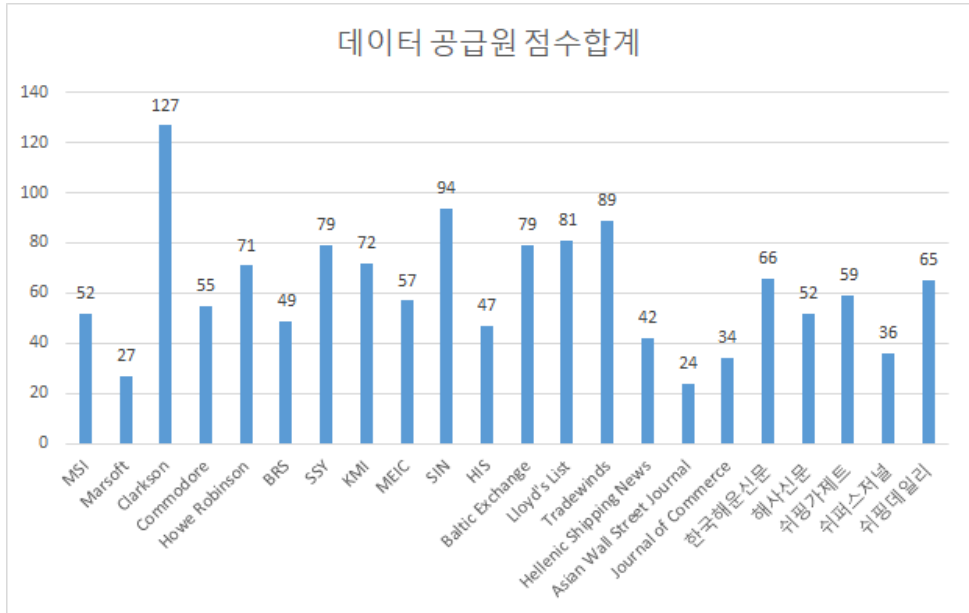
〈표 2-3〉 데이터의 공급원천

전문기관의 예측, 전망보고서 • MSI(Maritime Strategies International) • Marsoft
Clarkson, Commodore, KMI 등으로부터 유료 또는 무료로 확보할 수 있는 시황분석 보고서 Clarkson • Commodore • Howe Robinson • BRS • SSY • KMI • MEIC
가공되지 않은 데이터(raw data) • SIN(Clarkson Shipping Intelligence Network) • IHS • Baltic Exchange
뉴스매체 • Lloyd's List • Tradewinds • Hellenic Shipping News • Asian Wall Street Journal • Journal of Commerce • 한국해운신문 • 해사신문 • 쉬핑가제트 • 쉬퍼스저널 • 쉬핑데일리

자료 : 설문지에 제시된 예시(저자 작성)

그림 2-5는 위의 유형과 항목에 대하여 응답자가 표시한 활용도에 따라 높음은 3점, 보통은 2점 낮음은 1점으로 환산하여 집계한 결과이다. 유형별 점수의 평균을 보면 예측·전망보고서가 40점, 시황분석보고서가 73점, 가공되지 않은 데이터가 73점, 뉴스매체가 55점으로 외부 기관의 시황분석보고서에 대한 의존도가 가장 높고 가공되지 않은 데이터를 확보하여 자체적으로 분석하는 형태의 활용이 그 뒤를 따르고 있음을 알 수 있다.

〈그림 2-5〉 데이터 공급원천별 점수합계



매체별로는 국내의 해운기업들이 Clarkson에 의존하는 정도가 매우 큼을 알 수 있다. 수위를 차지한 것이 Clarkson에서 유료로 제공하는 시황분석 보고서이고 2위를 Clarkson에서 가공되지 않은 형태로 제공하는 플랫폼인 SIN이 차지했다.

뉴스미디어 중에서 Tradewinds가 가장 활용도가 높은 매체로 선정된 것에 대해서는 향후 추가적인 조사가 필요하다. Tradewinds가 다소 비공식적인 (informal) 성격이 있고 분석적이라기 보다는 업계의 동향을 다루는 미디어로 인식되는데 활용도는 Lloyds List 보다 높은 것으로 파악되었기 때문이다.

국내 미디어의 경우 변별력이 크지 않으나 한국해운신문과 쉬핑데일리의 활용도가 상대적으로 높은 것으로 파악되었다. 해외 해운 미디어의 평균이 72점이고 국내 해운미디어²⁸⁾의 평균이 61점으로 해외미디어의 활용도가 더 높은 것을 알 수 있다. 이는 국내 매체와 해외매체의 전달속도가 영향을 미친 결과로

28) 컨테이너서비스의 일정공시의 성격이 강한 쉬핑가제트와 쉬퍼스저널을 제외한 평균

판단된다.

예시되지 않아 추가로 기술된 자료원으로는 Middle East Business Intelligence, 해외건설협회, Banchemo Costa, Reuters, Bloomberg 등이 있었다. 하지만 추가로 제시된 매체가 소수인 점에 비추어 해운기업이 의사결정에 활용하는 데이터의 범위가 대체로 예시의 범주를 넘지 않는 것으로 판단된다.

4. 데이터의 활용 현황

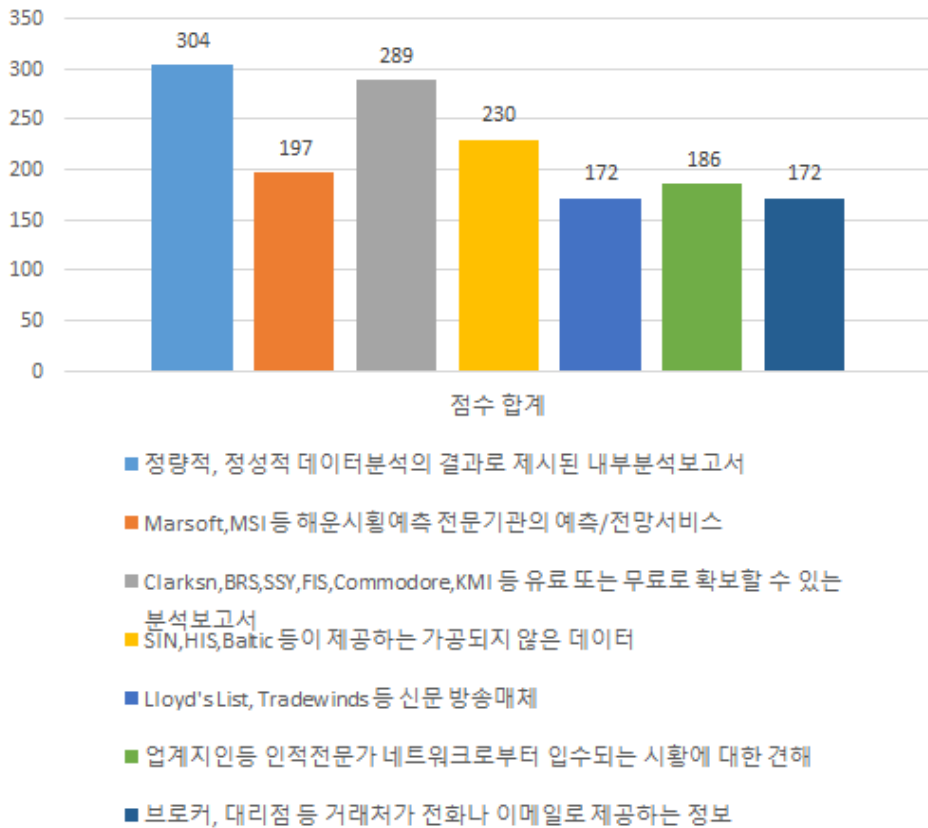
데이터는 의사결정의 준거로 활용된다. 설문조사에서는 의사결정 시에 해운기업이 활용하는 준거자료에 대하여 유형별로 순위를 파악하였다. 조사한 항목은 아래와 같다.

- 정량적, 정성적 데이터분석의 결과로 제시된 내부 분석 보고서
- Marsoft, MSI 등 해운시황 예측 전문기관의 예측/ 전망 서비스
- Clarkson, BRS, SSY, FIS, Commodore, KMI 등 유료 또는 무료로 확보할 수 있는 분석 보고서
- SIN, IHS Markit, Baltic Exchange 등이 제공하는 가공되지 않은 데이터
- Lloyds List, Tradewinds 등 신문, 방송매체
- 업계지인 등 인적전문가 네트워크로부터 입수되는 시황에 대한 견해(expert opinion)
- 브로커, 대리점 등 거래처가 전화나 이메일로 제공하는 정보

1순위 응답비율이 높았던 순서는 내부 분석보고서(28명), 외부 분석보고서(11명)이다. 내부분석 보고서에 대한 의존도가 높다는 것은 분석을 담당하는 인력이나 조직이 내부적으로 존재한다는 것이며 이들의 분석결과에 대한 신뢰가 있다는 것으로 해석할 수 있다. 외부의 분석보고서라고 응답한 비율은 전체의 20% 수준으로 외부 분석보고서를 그대로 의사결정에 반영하지는 않는 것으로 보인다.

순위를 점수로 환산해서 보면 다른 경로로 확보된 자료도 다양하게 의사결정의 준거로 활용하고 있음을 알 수 있다.

〈그림 2-6〉 데이터 활용현황 점수합계



제3절 전통적 시장분석의 현황과 한계

이 절에서는 국내의 해운기업들이 제 2절에서 파악된 전통적 데이터를 어떠한 기법으로 분석하는지, 분석은 실제 의사결정에 있어서 유용하다고 판단하는지, 분석결과를 실제 의사결정에 있어서 어떻게 활용하는지, 시장분석을 위해서

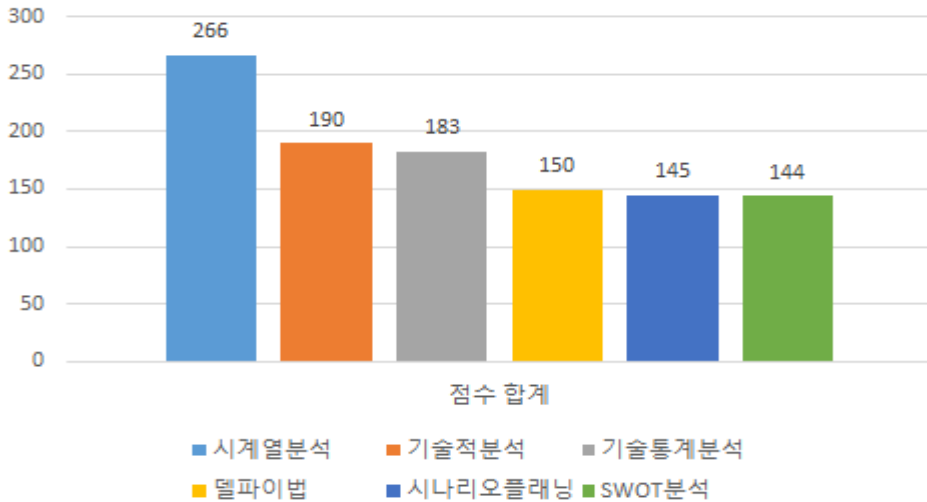
는 어떠한 조직과 인력을 운영하고 있는지를 살펴본다. 현황 분석은 2절에서와 마찬가지로 국내 주요 해운기업을 대상으로 한 설문조사의 응답(총 55건)을 토대로 분석 정리하였다.

1. 시장분석 기법

데이터에 근거하여 시장을 분석하는 데에는 2절에서 제시된 다양한 정량적, 정성적 기법들이 사용된다. 정량적 기법으로는 시계열 분석(time-series analysis), 기술적 분석(technical analysis)과 기술통계 분석(descriptive statistics)을 제시하였으며 정성적 분석으로는 델파이법(전문가의 주관적 의견의 종합), 시나리오 플래닝, SWOT분석을 제시하고 응답자에게 우선순위에 따른 순위를 부여하도록 요구하였다.

응답자들의 58%에 해당하는 32명이 시계열 분석의 활용도가 1순위라고 응답하였다. 순위를 점수로 환산하여 파악한 결과 시계열 분석, 기술적 분석, 기술통계분석으로 구성된 정량적 분석이 1~3위를 차지하여 정성적 분석보다는 정량적 분석을 선호하고 있는 것으로 확인되었다. 하지만 정성적 분석을 1순위로 선택한 응답자는 14명으로 전체의 25%를 차지하고 있고 3~5순위에 선택한 응답자가 많아 정성적 분석을 정량적 분석의 보조적 또는 추가적인 분석도구로 활용하고 있음을 알 수 있다. 아래 그림은 활용도 순위를 점수로 환산하여 표시한 것이다.

〈그림 2-7〉 시장분석기법의 활용도 순위



2. 정량적 분석의 유용성

정량적 분석을 주된 분석도구로 선택하는 것과 실제 활용에 있어서 유용성을 가지는 것은 별개의 문제이다. 이를 확인하기 위하여 다음과 같은 질문으로 정량적 시장분석의 유용성을 확인하였다.

- 계량적 시장분석은 의사결정에 도움이 되는 유용한 판단 기준이다.
- 계량적 시장분석은 의사결정의 참고자료 중 하나로 활용될 수 있다.
- 계량적 시장분석은 신뢰도가 낮아 의사결정에 거의 도움이 되지 않는다.
- 계량적 시장분석을 활용한 경험이 없어 잘 모르겠다.

본 설문은 중복응답을 허용한 관계로 누적비율이 100%를 초과하는 경우가 있음을 전제하고, 전체 응답의 76%에 해당하는 42명이 계량적 시장분석을 의사결정의 참고자료 중 하나로 활용될 수 있다고 응답하였다. 이는 1항의 시장분석 기법에 대한 현황이나 제 2절 전통적 데이터에 대한 현황분석 결과와 일치

하는 것이다. 즉 의사결정자는 다양한 원천으로부터 다양한 방법으로 정보를 받아들이며 의사결정의 정확도를 높이려고 노력하고 있다. 또한 이 결과는 Martin Stopford가 저서에서 밝힌 대로 시황예측의 역사적 성과가 미흡하여 의사결정자들이 계량적 예측의 결과를 크게 신뢰하지 않는다는 것을 뒷받침하는 것이다.

한편 계량분석이 신뢰도가 낮아 의사결정에 거의 도움이 되지 않는다는 응답자는 1명뿐으로 계량분석의 유용성을 부정하는 것이 아님이 확인되었다.

결론적으로 국내 해운업계는 계량적 시장분석을 전적으로 신뢰하지도 부정하지도 않으며 유용한 참고자료로 의사결정에 활용하고 있다.

3. 분석 결과의 활용

시장분석결과가 실제 의사결정과정에서 어떻게 활용되는지 확인하기 위하여 다음의 설문문항을 제시하였다.

시장분석담당자/조직의 의견이 실제 의사결정과정에서 활용된다.

- 시장분석담당자/조직의 의견이 잘 반영되는 편이다.
- 시장분석담당자/조직의 의견이 부분적으로 수용된다.
- 시장분석담당자/조직의 의견이 거의 받아들여지지 않는다.

시장분석담당자/조직의 의견이 활용되는 정도는 최고 의사결정권자의 리더십 성향에 따라 크게 변화한다.

첫 번째 문항은 3개의 하위 문항으로 구성되어 있으며 각 하위 문항은 상호 배타적이다. 하지만 첫 번째와 두 번째 문항은 상호 배타적이지 않다. 55명의 응답자 중 85%에 해당하는 47명의 응답자가 시장분석담당자/조직의 의견이 실제 의사결정에 활용된다고 답해 2절 4항에서 내부분석 보고서에 대한 의존이 크다고 분석되었던 것과 일치되는 결과를 보였다.

의견이 반영되는 정도에 대해서는 부분적으로 수용된다는 의견이 하위 항목 응답자 49명의 67%로서 다수를 차지했다. 이는 내부적인 분석이 이루어지더라도 의사결정권자의 의견이 적어도 동일한 비중을 차지하기 때문인 것으로 분석

된다.

분석결과 활용에 있어서 문제는 두 번째 항목이다. 시장분석담당자/조직의 의견이 활용되는 정도는 최고이사결정권자의 리더십 성향에 따라 크게 변화한다고 답한 응답자의 비율이 전체의 50%에 해당하여 기업의 사장이나 임원이 교체되었을 때 기업의 시장 분석기능에 큰 영향이 있을 수 있음을 시사하고 있다. 특히 대기업의 경우 최고경영자의 교체가 빈번하므로 그 영향이 클 수 있다. 이 결과는 다음 4항의 조직 분석에서도 확인되고 있다.

4. 시장분석을 위한 조직

시장분석 역량은 조직역량이어야 하며 연속성을 가지고 유지되어야 한다. 이를 확인하기 위하여 다음의 설문 문항을 제시하였다. 각 설문문항이 상호배타적이지 않으므로 응답자에게 복수의 응답을 허용하였다.

영업조직과는 별도로 시장분석을 위한 조직이 있다.

- 도의 시장분석 조직이 연속성을 가지고 운영된다(존폐가 반복되지 않는다)
- 시장분석조직의 구성원은 10명 이상이다.
- 시장분석조직의 구성원은 5명 이상 10명 미만이다.
- 시장분석조직의 구성원은 3명 이상 5명 미만이다.
- 시장분석 조직의 구성원은 1~2명이다.

영업조직에서 시장분석기능을 담당한다.

- 영업본부 내 별도의 담당자가 시장분석기능을 전담한다(영업기획 등)
- 영업팀장, 용선담당자 등이 시장분석을 담당한다.

시장분석기능이 체계적으로 이루어지지 않고 주로 경영층의 운영방침/ 지시에 따른다.

별도의 시장분석조직이 연속성을 가지고 운영된다(존폐가 반복되지 않는다)는

질문에 동의한 응답자는 25명으로 전체 응답자의 45%를 차지했다. 즉, 55%가 속한 집단에서는 시장분석조직 자체가 존폐를 반복한다는 것을 확인할 수가 있다. 이는 3항에서 최고 의사결정권자의 성향에 따라 시장분석조직의 의견이 활용되는 정도가 크게 변화한다는 분석과 일치하는 것이다. 이는 일본의 NYK나 MOL이 수십 년간 일관된 시장분석보고서를 발간하는 것이나 서양의 Clarkson Research, MSI(Maritime Strategies International) 등 전문 분석기관이 오랜 역사를 가지고 데이터 축적과 체계적인 분석을 진행하는 것과는 대조되는 것이다.

다음은 조직규모이다. 시장분석조직의 구성원에 대하여 구간을 10명 이상, 5~10명, 3~5명, 1~2명으로 나누어 조사한 결과 아래와 같은 빈도가 측정되었다.

〈표 2-4〉 시장분석 담당 조직규모

구간	빈도	비율
10명 이상	1	2%
5~10명	7	13%
3~5명	21	38%
1~2명	9	16%

위 표에서 보듯이 38%의 선사 또는 영업조직이 3~5명으로 시장분석조직을 운영하고 있는 것으로 파악되었다. 1~2명인 경우가 16%, 5~10명인 경우가 13%로 그 뒤를 따랐으며 10명 이상인 경우는 2%로 매우 낮았다.

영업조직에서 시장분석을 담당한다는 항목의 하위 질문으로 영업본부 내에 영업기획 등 별도의 담당자가 시장분석을 담당하는지 아니면 영업담당자가 시장분석기능을 겸하며 수행하는지 묻는 질문에 대하여 21명의 응답자가 영업본부 내 별도의 담당자가 전담하는 것으로 답했으며 29명은 영업담당자가 시장분석기능을 겸하는 것으로 응답했다.

영업담당자가 시장분석을 담당할 경우 단기 시황의 역동성이 효과적으로 반영되기 때문에 단기적인 의사결정에는 도움이 될 수 있으나 시간적인 자원배분이 요구되는 장기적이고 구조적인 예측에는 한계가 있어 효과적이지 않다.

마지막 질문은 경영층의 독단적인 의사결정이 이루어지는 것을 확인하는 것이다. 전체 응답자의 18%가 시장분석기능이 조직적/체계적으로 이루어지지 않고 주로 경영층의 운영방침/지시에 따른다고 응답하여 의사결정의 후진성을 드러내고 있다. 해운의 의사결정에 있어 고려해야 할 변수가 매우 복잡, 다양하고 와일드카드라고 할 수 있는 비체계적 이벤트들이 수시로 일어나기 때문에 경영층이 정보를 독점할 수 있는 가능성은 매우 낮다. 즉 조직역량이 체계적으로 갖추어지지 않으면 의사결정이 잘못될 가능성이 매우 커지는 것이다. 해운의 위기를 분석함에 있어 무능한 사주가 의사결정에 큰 영향을 미쳐 경영성고가 악화되는 소위 ‘오너리스크’가 거론되는 것도 이와 무관치 않다.

5. 전통적 시장분석의 한계

1) 데이터의 한계

원천자료가 거의 없는 국내 해운업계의 경우 소수의 상업적 공급기관이 제공하는 데이터에 의존하고 있는 것은 분석의 한계로 작용할 수 있다. 이들이 제공하는 데이터는 주로 시계열 데이터이며 전통적 데이터의 형태를 띠고 있어 4차 산업혁명이 가지고 온 새로운 데이터 수집방법을 제대로 수용하지 못하는 것도 한계이다.

2) 정량적 분석의 한계

계량분석 모델의 경우 한계를 보이는 경우가 많은데 M. Stopford²⁹⁾는 그의 저서 *Maritime Economics*에서 계량분석 모델이 문제가 되는 원인을 다음과 같이 열거하였다.

- 피상적 원인만 적시하는 부실한 모델설정
- 나무를 보고 숲은 보지 못하는 지엽적 구체화
- 상황에 따른 가정의 변화를 반영하지 못하고 당연할 것으로 생각되는 가정을 그대로 적용

²⁹⁾ Stopford, M. (2009), *Maritime Economics*, Routledge, Abingdon. p. 740

행동적 요소와 같이 예측이 불가능한 것을 예측하려는 시도

또한 정봉민(2015)³⁰⁾은 경기예측이 어려운 이유로 우발적 요인의 영향은 예측하기 어려운 한편 파급효과가 크고, 계량경제학적 모델이 미래는 과거의 반복이라는 전제하에 성립되는데 이는 현실과 괴리가 있으며, 행태적 요인이 예측의 결과에 영향을 미치고, 인간이 때로는 극점에서 ‘버블현상’과 같은 비이성적인 행위를 하기 때문으로 적시하였다.

3) 정성적 분석의 한계

정성적 분석의 경우 전문가의 의견을 종합하는 델파이법이나 시나리오 작성법 등을 활용하게 되는데 이 경우에도 여러 행태적 문제 때문에 분석에 한계를 초래하게 된다.³¹⁾

대부분의 사람들은 자기능력을 과신하거나 무능력하게 보이는 것을 회피하려는 성향을 가지고 있기 때문에 큰 범위로 모호하게 맞추기 보다는 정확하게 틀리는 것을 선택한다.

해운에서 두드러진 현상이기도 하지만 시황이 높을 때 높은 수준이 지속될 것으로 믿고, 낮으면 앞으로도 낮은 수준이 지속될 것으로 보는 현상유지성향을 갖는다.

위의 경우와 상반된 성향으로 어떤 부분에서는 현재로부터 파격적인 형태로 벗어날 것이라는 예측을 하기도 한다. 최근 McKinsey에서 50,000TEU급 메가 컨테이너선에 대해서 전망한 것도 실제 결과에 따라서는 그러한 예측의 사례가 될 수 있다.

주변의 압박(peer pressure)으로 같은 방향으로 보는 군중심리 상태에 놓이기도 한다.

시황예측기관들이 서로의 보고서를 참고하여 대체로 두드러져 보이지 않는 같은 결과를 내는 것과 같은 ‘눈치’가 작용하기도 한다.

30) 정봉민 (2015), 『해운경제학』, 블루&노트, pp. 209-210에 기초하여 저자 재정리

31) Stopford, M. (2009) *Maritime Economics*, Routledge, Abingdon. Ch.17에 기초하여 저자 재정리

이러한 일반적인 성향 이외에도 우리나라의 경우 강한 수직적 조직 문화로 정성적 예측을 시도할 때 고위 직급의 의견에 순응하는 형태의 편향이 존재하는 경우도 많이 관찰된다.

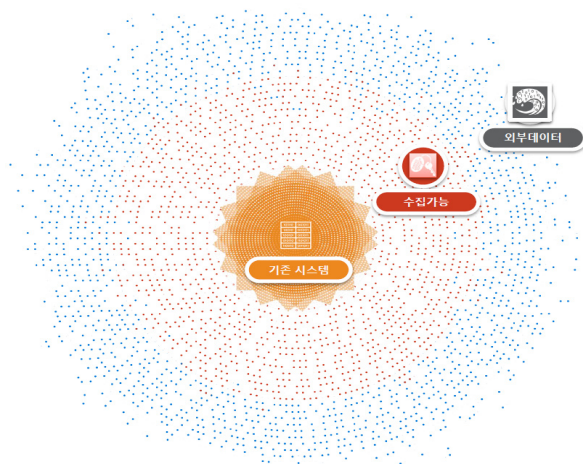
제3장 해운시장분석에 대한 새로운 접근법 – Big Data Analytics 《

제1절 빅데이터 분석

1. 개요

빅데이터에 대한 사회적 관심이 세계적으로 크게 고조되고 있다. 정보자산이란 그림 3-1과 같이 기존 시스템을 통해 수집·분석되고 있는 데이터와 수집 가능한 정보 및 그 밖에 외부데이터를 포함한 것으로, 빅데이터 분석은 이를 활용한 일관적인 분석을 통해 가치 있는 새로운 데이터를 추출하고 맞춤형 의사결정, 발견 및 프로세스 최적화를 향상시키기 위한 새로운 형태의 접근 방법이라 할 수 있다.

〈그림 3-1〉 데이터의 분포도



자료 : 장동인 (2015), CIO Summit 2015, http://www.cuvix.co.kr/cio_summit2015/Images/Track1-1_CIO_Summit_2015.pdf, 검색일 : 2017. 08. 21

미국 IT 자문기관 가트너(Gartner)에서는 유망 기술 중 하나로 빅데이터를 선정하였으며, “빅데이터란 높은 통찰력, 의사결정, 프로세스 자동화를 위해 비용효과가 높은 혁신적인 정보처리 과정을 요하며, 대용량의 데이터 규모 (high-volume), 빠른 속도(high-velocity), 높은 다양성(high-variety)을 지닌 정보 자산이다.”라고 정의하였다.³²⁾ 또한 미국은 빅데이터 관련 산업을 주도하기 위하여 관련연구 지원 및 빅데이터 기반의 기술 및 서비스 개발, 관련기관 협조, 데이터 관리, 인력 및 인프라 개발 등을 적극적으로 추진하고 있다. 미국의 빅데이터 관련 종사자 수는 2008년 150,000명에서 2018년 약 450,000명으로 급증할 것으로 예상된다.³³⁾

2. 등장 배경

정보통신 기술이 발전함에 따라 잠재되어 있는 새로운 형태의 정보자산으로 등장한 빅데이터는 대량의 정형·비정형 데이터 형태로 다양한 분야에서 매우 빠르게 수집되고 있다. 인터넷 검색사이트에서 모아지는 문서 및 사용자 정보, 소셜 네트워크 상의 사용자 관계정보, GPS 등 위치 정보, 글로벌 기상정보, 디지털 의료정보 및 유전자 정보 등 각 분야에서 모아지고 있는 데이터의 양은 기하급수적인 증가세를 보이고 있다. 이러한 데이터는 ‘21세기 원유’라고 표현될 정도로 매우 중요한 이슈로 다루어지고 있다.

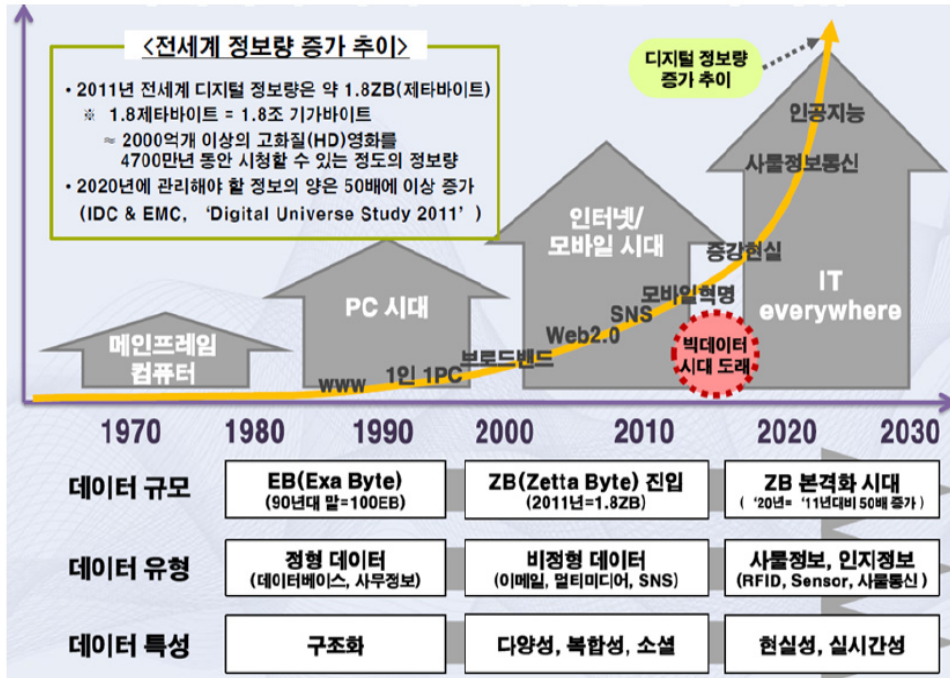
빅데이터는 그림 3-2와 같이 그 분량이 2020년 50배 이상 증가할 것으로 예상되고 있으며³⁴⁾, 응용 분야도 점차 다양화될 것으로 예상된다.

32) 임상노트 365, <http://kerneler.tistory.com/89>, 검색일: 2017.8.21

33) Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, A. H. (2011), *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute, New York. pp.103~106.

34) 김용대, 조광현(2013). 『빅데이터와 통계학』, 『한국데이터 정보학회지』 2013, pp. 959-974.

〈그림 3-2〉 빅데이터의 유형 및 규모



자료 : 김용대, 조광현(2013). 「빅데이터와 통계학」, 「한국데이터 정보학회지」 2013, p 960 재인용

우리나라의 경우 세계 최고 수준의 인프라를 기반으로 데이터의 생산 및 유통은 매년 급증하고 있으나, 기업의 데이터 활용 수준은 미흡했다. 이에 정부에서는 ‘스마트 국가 구현을 위한 빅데이터 마스터 플랜’을 수립하여, 여러 유관기관 간 협력 및 빅데이터 활용으로 국가 미래전략 지원, 공공서비스 혁신, 기업 경쟁력 제고, 생산성 향상 등을 도모하기 위한 계획을 추진하고 있다. 그러나 정부의 노력과는 달리 빅데이터 활용에 대한 국내 기업들의 인식이 부족한 것은 사실이다. 한계의 원인으로는 이해하기 어려운 복잡한 고급 통계와 수학적 기법을 사용하는 데이터 기반 의사결정으로의 전환에 대한 거부감, 빅데이터 분석 도입 효과에 대한 불신, 빅데이터 자체에 대한 낮은 이해도, 빅데이터 분석을 활용한 성공사례 부재, 국내 빅데이터 시장에 대한 정확한 시장 및 통계 자료의 부족 등이 거론되고 있다.³⁵⁾

3. 빅데이터의 중요성 및 가치

빅데이터와 분석 관련 기술은 소셜, 모바일, 클라우드와 더불어 디지털 시대의 변혁을 이끄는 주역으로 알려져 있다. 2016년 시장의 주인공이 BI(Business Intelligence) 강화를 주도한 빅데이터 기술이었다면 2017년은 데이터, 분석 분야의 혁신에 주목해야 할 한 해가 될 것이다³⁵⁾.

현재 빅데이터에 대한 관심이 급속도로 증가하고 있는 이유를 3가지로 설명할 수 있다. 첫째, 정보통신 기술 발달로 인하여 빅데이터가 출현했기 때문이다. 빅데이터는 일반적으로 데이터베이스에 저장되어 디지털화된 자료로, 이러한 대량의 자료가 지속적으로 수집되고 있다.

둘째, 빅데이터의 효율적인 저장 및 분석 기술이 어느 정도 가능해졌다. 과거에는 저장기술과 비용의 제약으로 대용량 데이터를 저장 및 관리, 분석하지 않았으나, 분산파일시스템(Distributed file system)의 등장으로 메모리 가격이 낮아지면서 저비용으로 대량의 자료의 저장이 가능하게 되었다. 또한 저장된 대량의 자료로부터의 정보 추출이 빠르게 이루어질 수 있으며, 특히 MapReduce를 사용하여 손쉽게 병렬처리가 가능해진 것이 주요인이라고 할 수 있다. 하둡(Hadoop)은 분산파일시스템과 MapReduce 프로그래밍 환경을 쉽게 사용할 수 있도록 구현한 시스템으로, 현재 빅데이터의 중요 기술로서 크게 각광을 받고 있다.

마지막으로 국가 및 회사 간 기술격차의 감소로 그 경쟁이 매우 치열해졌다. 기술격차가 감소함에 따라 새로운 작은 정보의 축적으로 매우 큰 영향을 미칠 수 있다.

업계는 빅데이터의 성장 잠재력과 가치에 주목하고 있다. 4차 산업혁명과 함께 대량의 데이터 생성이 급속도로 증가하고, 과거와는 다른 정보 유형 및 정보 소스를 통해 정보의 수집, 축적, 분석, 활용이 가능해졌다. 현재 시장에서는 이

³⁵⁾ 한국정보화진흥원 (2017), p. 14.

³⁶⁾ hor Olavsrud (2017), 전문가들이 말하는 2017년 빅데이터·분석 전망 15선,
<http://www.ciokorea.com/news/33014#csidxc14ca138a97d600b555c41c8c3f2d02>, 검색일 : 2017. 08. 21

를 활용하여 새로운 가치를 만들고자 하는 움직임이 나타나고 있으며, 이미 외국의 인터넷, 소프트웨어 기업 등뿐만 아니라 국내 시장에서도 비용 절감 및 효율성 개선, 마케팅 활용 등 빅데이터를 활용한 사례가 나타나고 있다.

우리나라 정부도 적극적인 빅데이터 활용을 위한 다양한 정책을 수립·시행하고 있다. 최근 정부 3.0 추진 기본계획을 발표하였고, 60개 이상의 정보 공유 및 시스템 연계·통합 과제와 빅데이터 시범과제를 추진하였으며, 과학기술기본계획에서 빅데이터를 30대 중점 기술로 선정하기도 하였다. 또한 미래창조과학부는 빅데이터를 활용하여 일자리, 청년 복지, 소상공인 지원, 보건의료 분야 등의 사회문제 해결을 위해 빅데이터 사업화 컨설팅 과제를 선정·지원하고 있다.

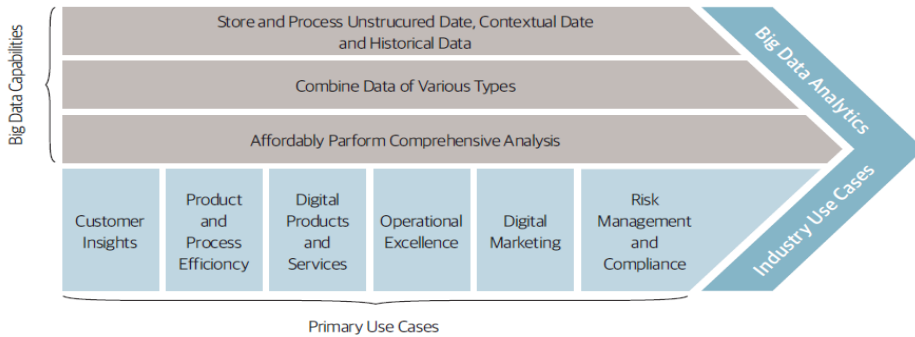
제2절 빅데이터의 정의 및 분석기법

1. 빅데이터의 정의

빅데이터(Big Data)는 기존 데이터베이스의 관리 도구를 활용한 데이터 수집, 저장, 관리, 분석을 위한 대량의 정형 및 비정형 데이터를 의미한다. 대규모의 다양한 데이터의 생성, 수집, 분석, 표현 등을 위한 빅데이터의 기술 발전은 다변화된 사회를 더욱 정확하고 효율적으로 예측하고 작동할 수 있도록 한다. 또한 개인화된 사회 구성원에 대하여 맞춤형 정보를 제공하고, 관리 및 분석이 가능하도록 하여 과거에는 불가능했던 기술을 실현시키기도 한다.

초기에는 빅데이터에 대한 기술적인 측면의 정의에서 출발하였으나 빅데이터의 가치와 활용효과 측면으로 그 의미가 확대되어 빅데이터를 단순한 정량적인 차원을 넘어 접근하고 있는 추세이다. 가트너그룹은 'The Big Data Value Model(2015)'을 통해 아래 그림과 같이 빅데이터 분석의 주요 활용 목적을 Customer Insight, Product and Process Efficiency, Digital Products and Service, Operational Excellence, Digital Marketing, Risk Management and Compliance의 여섯 개로 구분했다.

〈그림 3-3〉 가트너그룹의 'The Big Data Value Model'



자료 : 최재경, 「빅데이터 분석의 국내외 활용현황과 시사점」, KISTEP InI, 2016, p. 34 재인용

2. 빅데이터의 특징

더그 레이니(Doug Laney) 연구보고서(2001년)에서는 데이터의 급성장에 따른 이슈와 기회를 데이터의 규모(volume), 속도(velocity), 다양성(variety)이라는 세 개로 정의하고, 이후 “3V” 모델은 가장 일반적으로 사용되는 빅데이터의 정의가 되었다.³⁷⁾

〈그림 3-4〉 빅데이터의 4가지 차원



자료 : IBM 기업가치연구소 (2012), p. 4.

37) 위키백과, <https://ko.wikipedia.org>, (검색일: 2017.8.21)

빅데이터 기술이 발전하면서 정보 수집뿐만 아니라 적절한 시기에 필요한 정보를 추출하여 처리하는 것이 중요한 시대가 되었으며, 인공지능의 데이터 분석 능력과 대용량의 빅데이터를 처리할 수 있는 기계학습 알고리즘에 대한 관심이 갈수록 높아지고 있다. 또한 이미지와 음성 외에도 영상인식 등 다양한 형태의 고품질 데이터를 효율적으로 제공할 것으로 기대를 모으고 있는 딥러닝의 등장은 인공지능에 대한 새로운 패러다임이 되고 있다.

〈표 3-1〉 빅데이터의 특징

구분	내용
양 (Volume)	<ul style="list-style-type: none"> • 시스템의 데이터 크기, 데이터 처리 용량, 축적된 양을 의미한다. • 기하급수적으로 데이터가 생성되고 이를 수집, 저장, 처리하기 위한 기술이 핵심 역량이 되기 때문에 데이터의 규모는 중요한 속성이 된다. • 빅데이터의 크기는 움직이는 표적으로 과거보다 데이터의 규모가 더욱 증가하였으며, 여러 개발 요소들의 대규모 데이터(raw data, source data)의 집합이다.
속도 (Velocity)	<ul style="list-style-type: none"> • 빠른 데이터 생산 및 유통 속도, 이에 대한 실시간 분석과 처리의 중요성이 빅데이터의 속도적 특성이다. 즉 스트리밍, SNS 등 실시간성 데이터가 계속적으로 생성되고, 그 분석 결과를 실시간으로 활용하는 것을 추구한다. 또한 실시간적인 데이터 처리 및 분석 능력과 실시간적인 쌍방향 커뮤니케이션 능력을 모두 포함한다. • 취합된 데이터를 분석할 경우 과거에서 데이터가 생성된 시점에서 분석이 이루어져 시간차가 발생하게 되므로, 실시간 데이터의 경우 실시간 상황 분석이 특별한 의미를 가진다. • 데이터를 생성, 수집, 통합, 분석, 활용하는 모든 단계에서 속도가 중요하다.
다양성 (Variety)	<ul style="list-style-type: none"> • 다양한 종류의 데이터를 의미하며, 정형화의 종류에 따라 정형(Structured), 반정형(Semi-structured), 비정형(Unstructured)으로 구분한다. 즉 종류 및 형태가 다양한 데이터가 많은 채널로부터 수집되고, 구조화되지 않은 비정형 데이터가 증가됨을 의미한다. • 기존에는 내부의 관계형 데이터베이스 시스템에 저장된 정형데이터 분석이 주로 이루어졌다면, 최근에는 다양한 형태의 비정형 데이터에 대한 분석까지 이루어지면서 데이터가 다양해지게 되었다. • 빅데이터에서는 기존의 관계형 데이터뿐만 아니라 SNS, 위치정보, 멀티미디어 등 다양한 유형의 구조화되지 않은 데이터를 다룬다.
정확성 (Veracity)	<ul style="list-style-type: none"> • 데이터의 정확성은 해당 데이터에 대한 신뢰수준을 의미한다. • 높은 데이터 품질을 유지하는 것은 매우 중요하며, 최상의 데이터 정제(data cleansing) 기법 활용에도 불구하고 날씨, 경제, 고객의 미래 구매 결정 등 일부 본질적인 데이터의 불확실성은 제거할 수 없다.
가변성 (Variability) 복잡성	<ul style="list-style-type: none"> • 데이터 구조, 도메인 규칙, 저장타입 등 데이터의 생성, 처리, 가공 등의 과정 상 모든 요소가 복잡해지는 것을 의미한다. • 비정형 데이터 등 형식, 데이터 저장방식, 중복성, 외부 데이터의 유입 등 그 대상이 증가함

구분	내용
(Complexity)	<p>에 따라 데이터 처리 및 관리의 복잡성이 심화되고 있으며, 이를 위해 새로운 기법들을 요구하고 있다.</p> <ul style="list-style-type: none"> 데이터 처리의 복잡성은 타 영역과의 데이터 융합 또는 연계를 의미하며, 빅데이터의 가치를 극대화하기 위해서 분산되어 있는 수많은 데이터의 융합 및 연계 과정을 통한 분석이 필요하다.

자료 : 주해중·김혜선·김형로 (2017), pp. 34-36.

3. 빅데이터 분석단계

빅데이터 기술은 특정기준에 따라 아래 그림과 같이 적정한 데이터를 수집하고, 저장·분석하여 분석자 및 사용자들이 그 분석 결과를 쉽게 이해할 수 있도록 시각화하는 과정을 포함하며, 특히 다양한 형태의 고품질 데이터를 효율적으로 활용하는 데 있어 핵심 요소가 되는 시각화(visualization)는 매우 중요한 영역으로 주목받고 있다.

〈그림 3-5〉 빅데이터의 표현기술



자료 : 주해중·김혜선·김형로 (2017), p. 125.

1) 빅데이터 수집(Collection)

조직 내부 및 외부의 정형화된 데이터뿐만 아니라 다양한 형태의 데이터 중 필요로 하는 데이터를 선정하고, 이를 수집하기 위한 세부계획을 작성한 후 주

기적 또는 실시간으로 수집하는 과정으로 수집된 정보의 분석을 위하여 특정 데이터 형식으로 변환하는 것을 포함한다.

- 수집대상 데이터 선정 : 빅데이터 분석 및 서비스 제공을 위하여 품질을 결정하는 중요한 핵심 단계로 수집 대상 분야의 전문가 의견을 반영하여 분석 목적에 맞는 데이터를 선정하고, 수집가능 여부, 보안문제, 데이터의 정확성, 비용 등을 파악하여 세부 목록 및 항목을 작성함
- 수집 세부계획 수립 : 선정된 수집대상 데이터 유형을 분류하고, 데이터 소유자(내부/외부), 데이터의 유형, 수집 기술 및 주기 등을 포함한 세부계획을 수립함
 - (내부 데이터) 내부 시스템에서의 데이터 연계 가능 여부 등 파악
 - (외부 데이터) 개방 데이터의 종류 및 규모, 수집 시스템 연계방식 및 절차, 주기 등 관련 기술·정책을 파악
- 데이터 수집 실행 : 수집 계획에 따라 사전 테스트를 통해 관련 시스템을 점검한 후 실행하며, 데이터 수집 주체에 따라 능동적 또는 수동적 데이터 수집으로 분류함
 - 사전 테스트를 통해 보안성(개인정보 포함여부 등), 정확성(원본 데이터와의 비교), 네트워크상의 문제, 데이터 누락 등 점검
 - 데이터 수집 시 수집된 데이터의 수집 방법, 장애발생 여부, 시간, 출처, 등 수집 당시 상황 등을 체계적으로 기록
- 빅데이터 변환·통합 : 데이터수집 과정 중 컴퓨터가 바로 처리할 수 없는 비정형 데이터에 대하여 구조적 형태로 변환·통합·저장하는 과정으로, 비정형 데이터를 정제하거나 정형적 데이터의 누락, 오류, 형식차이 등을 제거함

2) 빅데이터 저장관리(Storage Processing)

빅데이터 저장관리는 다양한 형식의 대용량 수집 데이터 분석을 위해 적합한 방식으로 처리하여 데이터 품질을 향상시키고 안전하게 활용 할 수 있도록 변환, 통합, 처리하는 일련의 작업을 말한다.

- 빅데이터 전처리(Pre-processing) : 수집된 데이터를 저장하기 위한 단계로,

필터링을 거쳐 활용목적에 맞지 않는 정보를 제거하여 분석시간 단축 및 저장 공간의 최적화를 실현하고, 비정형 데이터의 경우 데이터 마이닝을 통해 중복, 오류, 저품질 데이터를 개선하는 과정을 말함

〈표 3-2〉 데이터 전처리 기술 고려사항

기술	고려사항
데이터 필터링 (Filtering)	<ul style="list-style-type: none"> • 데이터 필터링 기준을 설정할 수 있어야 함 • 데이터 처리 전후 생성된 파일에 대한 중복성 확인을 위해 파일명, 확장자 등 필터링 기능 제공 • 의미가 있는 데이터를 선별하기 위하여 먼저 정의된 기준을 비교 검증 할 수 있는 기능 제공 • 데이터 필터링 적용 시, 비정형 데이터 처리를 위하여 자연어처리 및 기계학습 전 사용자가 처리 방식을 선택하고 데이터에 대한 정형화기준 제공 • 수집된 데이터의 품질 기준 부합 여부 및 오류 등을 확인하고 관리자에게 알릴 수 있는 기능 구현 • 필터링 처리 시 사전에 정의된 기준에 의거하여 데이터 처리 중 발생된 오류에 대한 이력 저장기능 제공
데이터 유형 변환 (Transformation)	<ul style="list-style-type: none"> • 수집된 데이터에 대하여 유형 분류 시 적용할 수 있는 분류 기준 제공 • 데이터 유형을 분류하고, 데이터 변환을 위해 필요한 알고리즘 함수 및 변환 구조를 정의하는 기능 제공 • 사용자가 지정한 변환 형식에 맞게 이루어졌는지 확인할 수 있는 기능 제공 • 변환된 데이터를 저장하는 기능 제공
데이터 정제 (Cleansing)	<ul style="list-style-type: none"> • 데이터 정제 유형을 사전에 정의하고 속성 값을 부여하는 기능과 사용자가 직접 스크립트를 작성하는 기능 제공 • 사용자가 설정한 정제 방법을 사전에 정의하여 자동으로 지정할 수 있는 기능 제공 • 결측치, 잡음 데이터 처리, 데이터 저장, 제거 대상 삭제, 처리, 확인 기능 제공 • 불일치된 데이터를 교정하기 위하여 표현형식, 코드체계, 단위 등을 자동으로 교정하는 자동 스크립팅 기능 제공

자료 : 미래창조과학부 (2014), pp. 33-34.

- 빅데이터 후처리(Post-processing) : 저장된 데이터를 분석하기 전 단계로, 저장된 데이터를 분석이 용이하도록 변환, 통합, 축소 등 기술을 활용하여 가공하는 과정을 말함
- 평활화(Smoothing) : 데이터 추세에 벗어나는 값을 변환하여 데이터로

부터 잡음을 제거하는 것

- 집계(Aggregation) : 다양한 차원의 방법을 통해 데이터를 요약함
- 일반화(Generalization) : 특정 구간에 분포하는 값으로 스케일을 변화
- 정규화(Normalization) : 데이터에 대한 z-스코어 정규화, 소수 스케일링, 최소/최대 정규화 등의 통계적 기법을 적용
- 속성 생성(Feature Construction) : 데이터의 통합을 위하여 새로운 속성 또는 특징을 만드는 방법

3) 빅데이터 처리분석(Analysis)³⁸⁾

빅데이터 분석을 위한 플랫폼을 활용하여 수집·저장된 데이터를 분석 및 처리하는 과정으로, 대용량 데이터를 한번에 전략적으로 처리 분석해야 한다.

- 빅데이터 일괄 처리 : 대표적인 기술로는 하둡(Hadoop)의 맵리듀스(Map Reduce), 마이크로소프트의 드라이애드(Dryad)가 있음
- 빅데이터 분석 : 빅데이터에서 의미 있는 지식을 추출하고, 이를 효율적인 의사결정에 활용하기 위하여 다양한 분석 방법 및 인프라가 필요함

4) 빅데이터 시각화(Visualization)

빅데이터 분석 데이터에서 가치 있는 결과를 도출하기 위하여 수집, 저장, 처리 이후 데이터의 분석 및 활용을 위한 것으로, 광범위하게 분산된 방대한 양의 자료를 분석하여 한눈에 볼 수 있도록 도표나 차트 등을 활용하여 정리하는 과정이다. 벤 프라이(Ben Fry)의 데이터시각화 7가지 단계가 활용된다.

- 획득(Acquire) : 디스크, 파일, 네트워크를 통해서 시각화하고자 하는 데이터를 수집하는 과정
- 분류(Parse) : 데이터의 의미를 고려하여 구조적으로 카테고리를 설정하는 과정
- 선별(Filter) : 획득(acquire) 및 분류(parse) 과정을 바탕으로 시각화의 대상이 되는 의미 있는 데이터만 남기고 제거하는 과정

³⁸⁾ 주해중·김혜선·김형로 (2017), pp. 49-52.

- 마이닝(Mine) : 통계학, 데이터 마이닝 등 분석기법을 이용하여 정보 추출 알고리즘을 도출해 내는 과정
- 표현(Represent) : 마이닝(mining) 과정을 통해 도출해낸 결과를 효율적으로 표현할 수 있도록 막대그래프, 리스트(list)나 트리구조 등 기본적인 시각화 모델을 선택하는 과정
- 정제(Refine) : 기본적인 시각화 모델을 바탕으로 기본 표상(basic representation)을 보다 명확하게 정제하는 과정
- 상호작용(Interact) : 다양한 시각에서 시뮬레이션 할 수 있는 방법을 반영하는 과정

〈그림 3-6〉 빅데이터의 시각화 사례



자료 : ㈜위세아이텍, 「시각화를 활용한 빅데이터 분석과 공공데이터 개방사례」, 발표자료 활용, 2014.

5) 폐기(Disposition)

빅데이터 분석에 사용된 데이터에 대하여 삭제 및 폐기하는 단계로, 특히 개인 정보 및 정보의 가치가 없는 데이터들은 이용목적을 달성한 후에는 폐기하여야 한다.

4. 빅데이터 분석기법

최근 빅데이터 트렌드는 정보의 규모도 중요한 부분이지만 분석 및 활용 측면을 중요시 하고 있으며, 데이터의 홍수 속에서 빅데이터가 가지는 의미를 분석하고, 분석 결과를 바탕으로 새로운 가치 창출이 어떻게 이루어지느냐가 매우 중요하게 다루어진다. 즉, 실시간 빅데이터 분석 정보를 통해 실질적인 가치를 발견함으로써 빅데이터를 문제해결에 더 효과적으로 활용할 수 있다.

빅데이터 분석은 대량의 데이터로부터 숨겨진 패턴과 알려지지 않은 정보 간의 관계를 찾아내기 위한 과정이다. 데이터 분석에 사용하는 기술은 대부분 통계학과 전산학, 데이터 마이닝 분야에서 이미 사용하고 있으며, 이 분석 기술의 알고리즘을 대규모 데이터 처리에 맞게 개선하여 빅데이터 처리에 적용시키고 있는 것이다. 빅데이터의 형태(정형 및 비정형 빅데이터)와 해결하고자 하는 방법에 따라서 적용할 수 있는 분석 방법과 절차가 다르다. 아래에서는 데이터의 형태와 해결하고자 하는 방법에 따라 분석의 절차와 방법을 서술한다.³⁹⁾

1) 빅데이터 형태에 따른 분석 방법

(1) 정형 빅데이터를 위한 분석 기법

빅데이터를 포함하여 일반적인 정형 데이터를 분석하기 위한 방법론으로는 데이터 마이닝과 프로세스 마이닝이 있다.

데이터 마이닝을 이용하여 정형 빅데이터를 분석하는 방법으로는 대표적으로 SEMMA방법론, CRISP-DM방법론(Cross-Industry Standard Process for Data Mining), KDD방법론(Knowledge Discovery in Data base)등이 있으며 여기서는 대표적으로 많이 사용되는 SEMMA방법론 <그림 3-7>을 소개한다.

SEMMA방법론은 다음과 같은 단계로 수행된다.

- 데이터 추출(Sampling)
 - 분석 데이터를 생성

³⁹⁾ Joyce Jackson (March. 2002), "Data Mining: A conceptual overview"

- 데이터 전처리를 통한 분석 데이터 준비
- 데이터 탐색(Explore)
 - 기초 통계 및 시각화를 통한 데이터 탐색 및 이해
 - 검증하고자 하는 종속 변수를 선정
 - 변수 간 유의성 분석 및 상관 분석 수행
- 데이터 수정(Modify)
 - 분석 데이터의 수정 및 변환 : 수량화, 표준화, 정규화 등
 - 군집화 및 차원 축소를 통한 변수 변환
 - 군집화 : K-means, 계층적 군집, 비계층적 군집, 기댓값 최대화 등
 - 차원 축소 : 주성분 분석, 특이값 분해, 비음수 행렬 인수분해 등
 - 최적의 모델을 구축할 수 있도록 다양한 형태로 변수를 생성, 선택, 변형 수행
- 모델링(Modeling)
 - 변수 형태(독립/종속)에 맞는 모델링 방법 설정

〈표 3-3〉 변수 형태(독립/종속)별 모델링 방법

	연속형 독립변수	범주형 독립변수
연속형 종속변수	Linear Regression Neural Networks K-Nearest Neighbors	Logistic Regression Linear Discriminat Analysis K-Nearest Neighbors
범주형 종속변수	Linear Regression Neural Networks Regression Tree	Logistic Regression Neural Networks Decision Tree Naive Bayes

자료 : Kuhn, Max, and Kjell Johnson. Applied predictive modeling. Vol. 810. New York: Springer, 2013.의 내용 전반 저자 정리

- 활용 목적(분류/예측)에 따른 모델링 방법 설정

〈표 3-4〉 활용 목적(분류/예측)별 모델링 방법

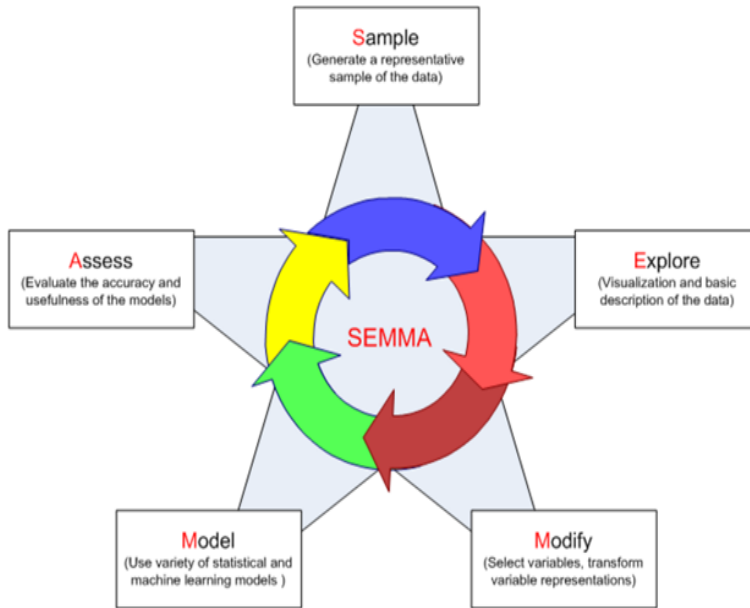
	예측	분류
선형	Linear Regression Partial Least Squares (PLS) Penalized Models	Logistic Regression Linear Discriminant Analysis PLS Discriminant Analysis
비선형	Neural Networks Support Vector Machines K-Nearest Neighbors	Neural Networks Support Vector Machines K-Nearest Neighbors Naive Bayes
규칙 기반	Regression Tree Random Forest Bagging	Decision Tree Rule-Base Models Random Forests Bagging

자료 : Kuhn, Max, and Kjell Johnson. Applied predictive modeling. Vol. 810. New York: Springer, 2013.의 내용 전반 저자 정리

○ 모델 평가(Assessment)

- 모델링 단계에서 추출 된 모델과 알고리즘을 평가하고 선택
- 모델의 평가 방법
 - 분류 모형 : Accuracy, Error Rate, Sensitivity, False Positive Rate, Precision, Specificity, False Positive Rate, F-Score, ROC Curve, Recall, Precision Recall Plot
 - 회귀 모형 : Root Mean Squared Error, Mean Squared Error, Mean Absolute Error, Relative Absolute Error, Relative Squared Error, Coefficient of Determination
- 모델의 알고리즘 선택 방법: Cross validation, Hold-out cross validation, K-fold cross validation, Leave-one-out cross validation

〈그림 3-7〉 SEMMA 방법론을 활용한 데이터 마이닝 절차

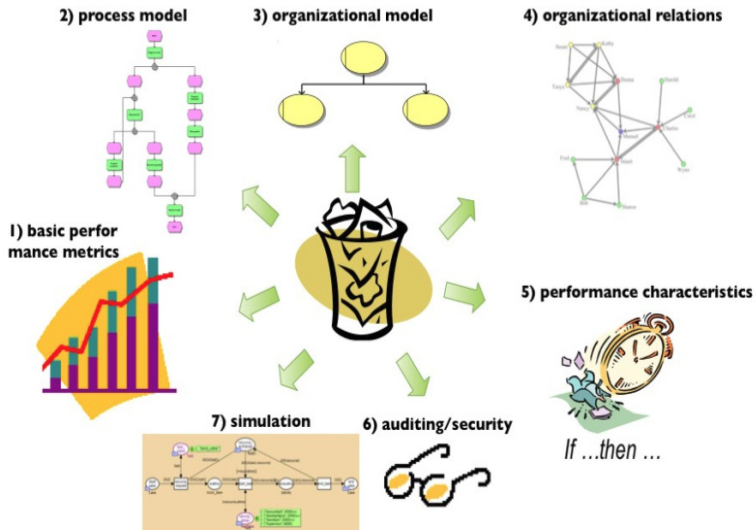


자료 : Data Mining Using SAS® Enterprise Miner™: A Case Study Approach, Third Edition

일반적으로 데이터 마이닝은 위에서 설명한 SEMMA방법을 적용하여 정형 데이터로부터 분석을 수행한다. 각 해당 단계에서 상황과 목적에 맞는 적절한 분석 방법을 선택해야 보다 유의미한 정보를 추출할 수 있으며, 해결하고자 하는 문제에 맞는 모델링 기법을 잘 선택하여야 정형 빅데이터로부터 효과적인 해결 방법을 도출해 낼 수 있다. 하지만 프로세스라는 관점에서 데이터 마이닝 기법을 활용할 경우는 분석 수행에 한계가 존재한다. 프로세스 관점에서 분석을 수행하고 이를 개선시키기 위한 방법으로는 프로세스 마이닝이 존재한다.

프로세스 마이닝은 BPM, ERP, MES, CRM 등 다양한 기업의 업무 처리 시스템에서 기록되는 빅데이터의 이벤트 로그를 분석하여 프로세스 관점에서 의미 있는 정보를 찾아내는 것을 목적으로 한다. 아래 〈그림 3-8〉는 프로세스 마이닝의 세부적인 연구 분야들을 나타낸 그림이며 프로세스 모델을 도출하고, 조직 관점에서 프로세스를 분석하고 성능을 평가하고 분석된 결과를 이용하여 시뮬레이션을 수행함으로써 프로세스를 개선하고 최적화 하는데 활용될 수 있다.⁴⁰⁾

〈그림 3-8〉 프로세스 마이닝의 연구 분야



자료 : 한국 방송 통신 전파 진흥원, “빅데이터 분석을 위한 프로세스 마이닝 기술 동향”, 2014, p. 3.

(2) 비·반정형 데이터를 위한 분석 기법⁴¹⁾

비·반정형 빅데이터에 해당하는 문서, 영상, 음성 및 GPS, AIS와 같은 데이터들은 정형의 데이터 분석을 위해서 사용되는 데이터 마이닝과 같은 방법으로는 분석이 불가능하기 때문에 이러한 데이터를 분석하기 위한 방법들이 활발히 연구되고 있다. 비·반정형 빅데이터를 분석하기 위한 방법으로는 문서 영상, 음성 등을 처리하기 위한 텍스트 마이닝과 GPS, AIS등을 처리하기 위한 트라젝토리 마이닝이 존재한다.

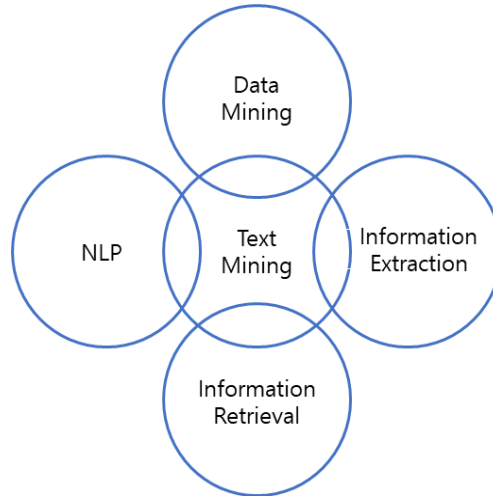
텍스트 마이닝은 〈그림3-9〉과 같이 자연어 처리 기술(NLP)과 정보 검색(information retrieval), 정보 추출(information extraction)방법 그리고 기존의 데이터 마이닝과 결합하여 비·반정형 빅데이터로부터 유용한 정보를 추출하고, 가공하는 목적으로 사용된다. 대용량의 텍스트 데이터에서 의미 있는 정

40) 배혜림 (2015), “프로세스 마이닝을 이용한 생산 운영 빅데이터 분석 동향 및 사례”, 『e 매거진』, 제 22권 제 2호, pp. 36-42.

41) Jeffrey L. Solka (2008), “Text Data Mining: Theory and Methods”

보를 추출하고, 해당 정보와 연계된 정보를 파악하는 등 텍스트가 가진 단순한 정보 이상의 의미 있는 결과를 얻어 낼 수 있다는 장점을 가지고 있다.

〈그림 3-9〉 Text Mining Areas



자료 : G.Miner, D.Delen, J.Elder, A.Fast, T.Hill, R.Nisbet, Elsevier (2012), "Practical text mining and statistical analysis for non-structured text data applications", chap. 2. p.31

텍스트 마이닝은 아래와 같이 5단계로 수행이 된다.

- 텍스트 전처리(Text Preprocessing)
 - Text cleanup: 바이너리 형식으로 변환 된 텍스트를 정규화하거나, 표, 그림 및 수식을 처리하는 등 불필요하거나 원치 않는 정보를 제거하는 것을 의미
 - Tokenization: 토큰 화는 앞의 단계에서 확인 된 약어에 속하지 않는 문장 부호 및 공백에 텍스트를 분할
 - Part of speech tagging: 문장 내 단어들의 품사를 식별하여 태그를 붙여주는 것을 의미
- 텍스트 변환(Text Transformation)
 - 문서로부터 추출된 단어들의 관계 및 분포를 고려하여 단어의 집합을 저장하고 관리하는 단계

- Vector space model, Bag of words 등의 방법을 이용해서 문서를 벡터로 공간에 표현하는 단계
- 특징 선택(Feature Selection)
 - Data mining에서 변수 선택(modify)에 해당하는 단계
 - Information gain, chi-square statistic, Entropy-based ranking, Document frequency, Term strength, 등의 방법들이 존재
 - 위의 특징 선택 방법들은 여러 번의 반복 수행과 적절한 방법의 사용을 통하여 최적의 특징을 선택하며, 분석 결과 및 모델의 품질은 선택된 특징에 종속됨
- 분석(Analysis) 및 모델링(Modeling)
 - Clustering: 단어가 가지는 유사성을 평가한 후 비슷한 성격을 가지는 단어들끼리 묶는 분석 방법으로, Probabilistic clustering algorithm, k-means clustering algorithm, Hierarchical clustering algorithm 등의 방법 이 존재
 - Social network analysis : 텍스트 간의 관계를 노드와 링크로 모델링 하는 방법으로 텍스트 간의 숨겨진 관계를 표현하고 도출 하는 방법
 - Sentiment analysis : 언어에 나타난 주관성 요소를 탐지하며 단어가 가지는 긍정 및 부정의 정도를 판별하고 정량화하는 분석 방법으로 Knowledge-based method, Statistical method 등의 방법이 존재
- 모델링(Modeling)
 - Classification: Naive Bayes Classifier, Nearest Neighbor Classifier, Support Vector Machines, Decision Tree classifiers 등이 존재
 - Topic model and Question answering model: Latent Dirichlet Allocation, Probabilistic Latent Semantic Analysis 등이 존재

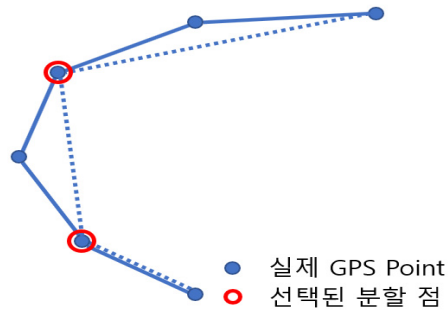
GPS, AIS데이터의 경우는 문서나 음성, 영상과는 다른 형태의 비·반정형 데이터로서 이동하는 특정 사물의 특정 시점에서의 위치를 연속적으로 기록한 데이터이다. 이러한 데이터를 분석하기 위해서 트라젝토리 마이닝이라는 분석 방법을 이용하며 수집된 이동 신호들을 군집화하고 이들의 분류를 통해 이동 경로를 분석하고, 더 나아가 이동 경로를 예측한다. 트라젝토리 마이닝은 다음과 같

이 두 가지 절차로 수행한다.⁴²⁾

○ Trajectory preprocessing(신호 전처리)

- GPS, AIS로부터 수집된 데이터를 분석 할 수 있도록 수치화 및 수량화를 통해 전처리하는 단계
- 분할 : 개별 위치가 수집된 점 데이터를 선분으로 나누는 과정으로 <그림 3-10>과 같이 모든 점들 간의 오차가 최소가 되며, 최소한의 점을 연결하는 선분을 구하는 분할점을 선택하여 데이터들을 선분으로 표현하는 단계

<그림 3-10> GPS신호의 분할



- 군집화 : 분할을 통해 표현된 선분들을 대상으로 군집화를 수행하며, Density-based method, Statistical-based method 등의 방법들을 통해 군집화를 수행하여 전체 신호를 대표할 수 있는 궤적들로 표현하는 단계
- 분류 : 군집화를 통해 표현된 궤적 속에서 위치 기반의 특성 및 공통 궤적 기반의 특성을 추출 하고 분류하는 단계로 SVM과 같은 방법들을 이용하여 특성을 분류

○ Trajectory Analysis

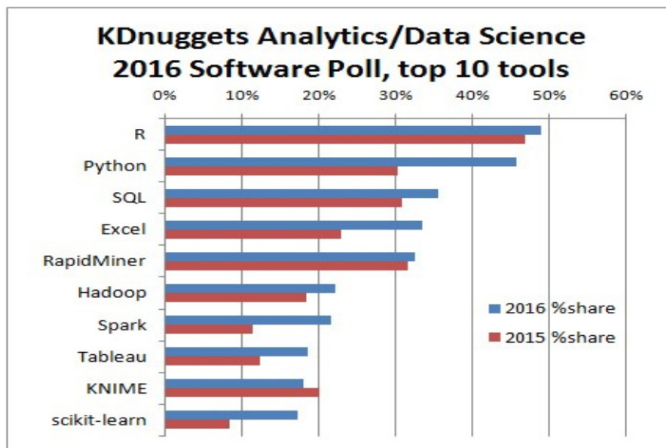
⁴²⁾ Feng Z, Zhu Y (2016), "A Survey on Trajectory data Mining Techniques and Applications", IEEE Access

- 패턴 분석 : 단순 빈도를 기반으로 한 패턴 분석, 이동 객체를 그룹화하여 각 이동 객체의 빈도를 이용한 패턴 분석 등을 통해 유의미한 이동 패턴을 추출
- 이상치 탐색 : 패턴 분석이 주요한 이동 궤적 패턴을 추출하는 것이라면, 이상치 탐색은 패턴 분석을 기반으로 하여 대다수의 이동 궤적과 멀리 떨어져 있는 이동 패턴을 발견함으로써 이상 이동 패턴을 발견하는 방법을 의미
- 예측 : 분석 결과를 통해 추출된 주요 이동 패턴과 이상 이동 패턴을 이용하여 움직이는 물체의 이동 패턴을 예측하는데 활용되며 Markov probability model이나 sequential rule based model, deep learning 등의 방법론을 활용하여 예측을 수행

2) 빅데이터 분석을 위한 툴

다양한 정형 및 비·반정형 빅데이터를 분석하기 위해서는 분석 방법론만큼 분석에 활용이 되는 툴도 중요하다.

〈그림 3-11〉 2016년 빅데이터 분석 툴 사용설문 조사 결과



자료 : R, Python Duel As Top Analytics, Data Science software - KDnuggets 2016 Software Poll Results, www.kdnuggets.com/2016/06/r-python-top-analytics-data-mining-data-science-software.html

빅데이터 분석을 위한 툴은 다양하나 그중에서도 R과 Python이 가장 많이 사용되고 있는 것으로 알려져 있다. R과 Python의 장단점은 표<3-5>와 같으며 R은 연구 기관이나 분석 전문가들이 주로 사용하고, Python은 개발자나 프로그래머들이 주로 사용하고 있다.

〈표 3-5〉 R과 Python의 비교

	R	Python
개발 목적	데이터 분석 및 통계 분석, 그래프 모델에 특화되어 개발된 언어	생산성과 코드의 가독성을 목적으로 개발된 범용적인 언어
장점	활발한 커뮤니티 운영 방대한 패키지 제공 시각화 모델링에 특화	비정형 데이터 처리에 특화 많은 딥러닝 전용 라이브러리 제공
단점	상대적으로 배우기 어려움 상대적으로 무겁고 느림	R에 비해 적은 패키지 제공 - 시각화가 어려움
주요 사용자	통계 전문가, 금융 분석 전문가, 데이터 분석 전문가	데이터 분석 및 통계 기술 분야 개발자 및 프로그래머

자료 : KDnuggets, "R vs Python for Data Science: The Winner is ...",
www.kdnuggets.com/2015/05/r-vs-python-data-science.html

두 가지 툴 모두 빅데이터 분석에 적합하나 R은 데이터 전처리 및 기초 통계 분석 및 모델링 그리고 시각화에 장점이 있으며, Python은 텍스트 처리, 음성 처리와 같은 비정형 데이터 처리에 장점을 보유하고 있다. 더욱이 빅데이터와 함께 거론되는 주요 키워드중 하나인 딥러닝을 사용하기 위해서는 Python에서 제공하는 tensorflow가 필수적인 도구로 활용되고 있다. 최근에는 R에서도 GPU를 이용한 tensorflow가 사용되고, 딥러닝 모델을 포괄하려는 노력들과 함께 R을 활용한 딥러닝 모델의 시각화 및 학습 과정 애니메이션화의 사례들도 늘고 있는 추세이지만 딥러닝 관련 분야에서는 Python이 여전히 우위를 점하고 있는 것으로 알려져 있다.

제3절 빅데이터 분석에 대한 연구

1. 해운에 대한 연구

해운산업에서는 선박의 운영, 오염물질 배출량, 에너지 소비 및 유지·보수 매커니즘 등을 보다 잘 이해하고 개선하기 위한 빅데이터 활용에 관심이 높아지고 있다. 하지만 타 분야에 비해 상대적으로 세부 주제별 빅데이터 수집 및 분석 방법론에 대한 학술적인 근거가 부족한 상황이다. 다만 빅데이터의 범위를 해운 활동에서 파생되는 모든 활용 가능한 데이터로 확장할 경우에는 보다 다양한 연구가 진행 중이다.

Rodseth(2016)⁴³⁾는 선박 운항에서 수집 가능한 원천 데이터의 종류와 데이터 수집 및 처리 시 발생하는 문제점을 정리하였다. 또한 발생한 문제점을 해결하고 e-Navigation을 포함한 데이터 수집 및 처리 프로세스를 단순화시키는 새로운 표준 프로토콜을 제시하였다. 아래 <표 3-6>은 선박 운항 시 발생하는 데이터의 종류와 각 데이터를 수집·분석할 때 주의해야 할 사항을 요약하고 있다.

43) Rodseth, O. J., L. P. Perera and B. Mo, "Big Data in Shipping – Challenges and Opportunities", In proceedings of the 15th International Conference on Computer Applications and Information Technology in the Maritime Industries (COMPIT 2016), 2016

〈표 3-6〉 데이터 종류와 수집 시 문제점

데이터 종류 문제점	1° Bridge	2° Automation	3° CPS	4° Performance	5° Reporting	6° Monitoring	7° Weather	8° Port data
컨텍스트 의존성 문제	X	X	X					
데이터 보안 문제	X	X	X					
복합적 현상 측정의 어려움	X			X			X	
비자동 데이터 입력 오류					X	X		X
상업적 이슈에 의한 고의적 오류					X			X
각종 센서 등 기기의 품질문제		X						
독점적이며 비싼 데이터 인터페이스		X	X					
파생 데이터의 소유권 문제			X	X				
인터페이스 표준화 문제		X	X	X				
AIS 기지국과 위성 간 불규칙한 데이터 수집						X		
사이버 공격	X				X	X		

자료 : Rodseth(2016)⁴⁴⁾의 내용을 저자가 번역하여 재작성

Lun(2009)⁴⁵⁾은 선사의 입장에서 가장 비용 효율적인 선단 구성에 대한 연구 결과를 발표하였다. 논문에 따르면 비용 효율성 달성을 위해 선박의 대형화가 진행되는 가운데 실제 더 나은 기업 실적을 달성하기 위해서는 해운 서비스와 선박의 크기 사이에 균형이 필요하다. 경로분석(path analysis)을 통해 선박의 수와 평균 크기가 해운회사의 실적에 미치는 영향을 분석하였으며, 또한 해운 서비스 빈도, 고객 가치, 최적 선박 크기, 기항지와 광범위 시장 범위 등으로 구성

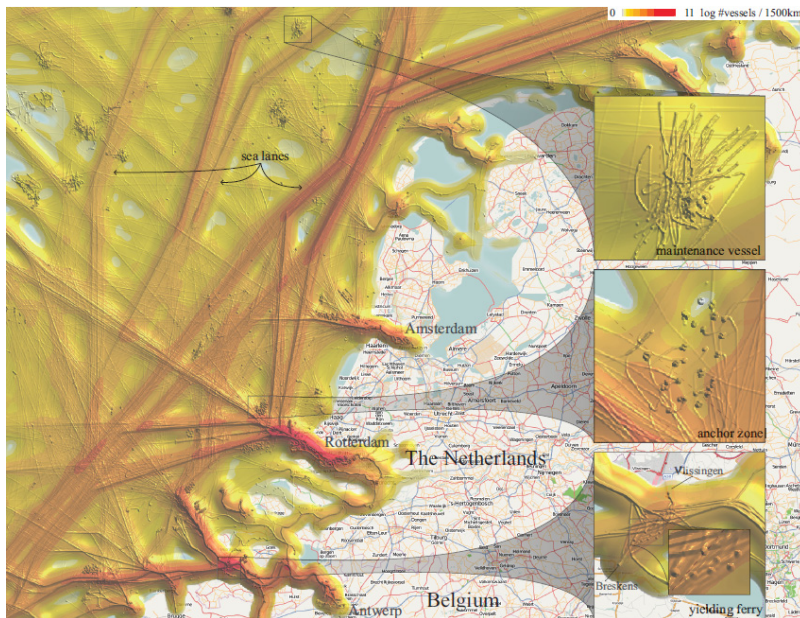
⁴⁴⁾ Rodseth, O. J., L. P. Perera and B. Mo, "Big Data in Shipping - Challenges and Opportunities", In proceedings of the 15th International Conference on Computer Applications and Information Technology in the Maritime Industries (COMPIT 2016), 2016

⁴⁵⁾ Lun, V. H. and M. Browne, "Fleet Mix in Container Shipping Operations", *International Journal of Shipping and Transport Logistics*, Vol. 1, No. 2, 2009

된 ‘SCOPE’ 프레임워크를 소개하였다. 이 프레임워크는 선사가 정기선 운용에서 최적의 선단 구성하도록 돕는 것을 목적으로 하고 있다.

Willems(2009)⁴⁶⁾는 AIS 데이터를 활용하여 항해 중인 선박의 위치에 대한 지리적 시각화 방법론을 제시하였다. 이 방법론은 선박이 미확인 지역에서 항로 및 정박 구역과 같은 중요한 해상 지역을 식별하는 데 도움을 준다. 시각화는 동적인 선박 위치에 커널을 중첩시켜 생성된 밀도장(density fields)에 기반하여 구축되며, 색깔로 구분되는 높이 지도에 표현된다. 이 시각화 방법론은 모든 움직이는 물체 데이터의 시각화에 적용될 수 있다.

〈그림 3-12〉 지리적 시각화 예시(네덜란드 연안)



자료 : Willems, N., H. V. D. Wetering and J. V. Wijk, "Visualization of Vessel Movements", *Computer Graphics Forum*, Vol. 28, No. 3, p.7, 2009.

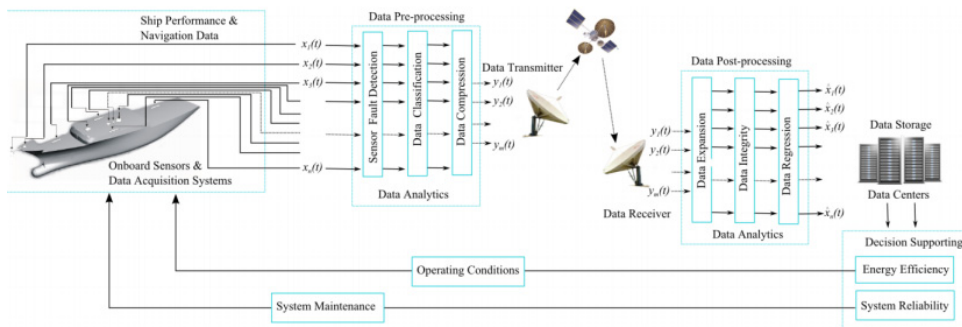
Perera(2016)⁴⁷⁾는 선박의 성능 및 항법 데이터를 활용하여 선박에너지효율

⁴⁶⁾ Willems, N., H. V. D. Wetering and J. V. Wijk, "Visualization of Vessel Movements", *Computer Graphics Forum*, Vol. 28, No. 3, 2009

⁴⁷⁾ Perera, L. P. and B. Mo, "Machine Intelligence for Energy Efficient Ships : A Big Data Solution",

관리계획(SEEMP : Ship Energy Efficiency Management Plan)의 일환으로 적합한 항법 전략을 개발하였다. 다양한 항법 정보를 갖춘 SEEMP는 통합선교 시스템(integrated bridge system)의 e-Navigation에서 중요한 역할을 한다. 선상 대규모 데이터 처리 프로세스를 통해 효과적인 항로 탐색이 필요한 상황에서 정확한 선박의 성능과 네비게이션 정보를 이용한다. Perera(2016)가 제안한 선박 성능 및 항법 데이터 모니터링을 위한 데이터 플로우차트(flow chart)는 빅데이터 솔루션으로 항법 전략의 질을 향상시킨다. 제안된 플로우차트는 아래 <그림 3-13>과 같이 사전 및 사후 프로세스로 나뉘며, 사전 프로세스는 선상 애플리케이션으로 센서 결합 탐지, 데이터 분류 및 데이터 압축으로 구성된다. 반면 사후 프로세스는 연안의 데이터센터에서 사용되는 애플리케이션으로 데이터의 확장, 무결성 검증 및 분석 단계로 구성된다. 각 단계에서는 다양한 머신러닝 방법론을 사용하여 빅데이터 분석을 수행한다.

<그림 3-13> 데이터 플로우차트



자료 : Perera, L. P. and B. Mo, "Machine Intelligence for Energy Efficient Ships : A Big Data Solution", *Maritime Technology and Engineering*, CRC press, 2016, p.143

해상교통 분야에도 다양한 데이터를 활용한 확률분석 모형이 사용되고 있다. Goerlandt(2011)⁴⁸⁾는 선박이 서로 충돌할 확률을 평가하는 방법을 제시하였

Maritime Technology and Engineering, CRC press, 2016

⁴⁸⁾ Goerlandt, F. and P. Kujala, "Traffic Simulation Based Ship Collision Probability Modeling", *Reliability Engineering & System Safety*, Vol. 96, No. 1, 2011

다. 선박의 충돌은 인간, 환경 및 경제적 손실 측면에서 다양한 위험을 초래하며, 이에 사고의 확률과 결과에 대한 합리적인 추정치를 산출하는 것이 중요하다. 제시된 방법론은 기대 사고 횟수와 사고가 발생할 가능성이 가장 높은 위치와 시간을 결정할 수 있으며, 사고에 따른 피해규모 예측모형 내 변수의 투입량을 제공한다. 또한 충돌감지 알고리즘을 이용하여 특정 지역에서 선박 트래픽에 대한 광범위 시간영역 마이크로 시뮬레이션을 수행한다. 시뮬레이션의 변수로는 각 항로의 선박 수, 선박 출발 시간, 선박의 크기 및 항해 속도 등이 이용되며, 투입량은 AIS 데이터에 기반하여 분석된다.

Stateczny(2014)⁴⁹⁾는 전자해도(ENC: Electronic Navigation Chart) 제작 시 사운드 정보 수집의 필요성에 따라 대용량 수역 데이터의 차원 축소 문제를 다루고 있다. 멀티빔 수중음파 탐지기의 변형인 간섭계(interferometric) 수중음파 탐지기를 통해 데이터를 수집하였으며, 보다 쉽고 효과적인 분석을 위해서는 데이터의 정리 및 축소가 필요하다. 이에 Stateczny(2014)는 인공신경망(ANN: Artificial Neural Networks)을 활용하여 데이터를 클러스터링하는 알고리즘을 제시하였다. 제안된 솔루션은 Kohonen 네트워크를 기반으로 하고 있으며, 최적 네트워크 구성에 대한 실험 결과를 제시하고 있다.

2. 타 분야의 연구

1) 물류와 공급망 관리(Logistics and Supply Chain Management)

해운산업과 직·간접적으로 연관된 산업 중 빅데이터 활용에 대한 연구가 활발한 분야로는 물류(logistics)와 공급망관리(SCM: Supply Chain Management) 등이 있다. SCM은 부품 제공업자로부터 생산자, 유통업자, 고객에 이르는 물류의 흐름을 하나의 가치 사슬 관점에서 파악하고 필요한 정보가 원활히 흐르도록 지원하는 시스템을 말한다.⁵⁰⁾ SCM에는 판매, 소비자, 재고관리, 위치 및 시간

49) Stateczny, A. and M. Włodarczyk-Sielicka, "Self-organizing Artificial Neural Networks into Hydrographic Big Data Reduction Process", *Rough Sets and Intelligent Systems Paradigms*, Chapter 34, Springer, 2014

50) 국가기술표준원, 『표준기반 R&D 로드맵 스마트SCM』, 2015, p.9

등 다양한 종류의 구조화 또는 비구조화된 대용량 데이터가 존재한다. Singh(2016)⁵¹⁾은 Delphi 기법을 사용하여 SCM의 주요 문제점을 도출하고 빅데이터 분석을 활용한 해결책을 모색하였다. 빅데이터 활용을 통해 달성 가능한 효과로는 물품 배송시간 감소, 고객 서비스 향상, 재고 물품 관리비 감소, 운영비 감소 등이 제시되었다.

Witkowski(2017)⁵²⁾는 기술과 조직 두 영역에서 IoT, 빅데이터 등 4차 산업혁명 기술을 활용한 혁신적 스마트 솔루션의 도입을 주장하였다. 특히 물류와 SCM에서 스마트 솔루션 도입은 개별 회사뿐만 아니라 국가 경제, 심지어 세계 경제에도 영향을 미치며, 이러한 혁신이 시장의 진전을 주도하게 될 것임을 지적하였다.

〈그림 3-14〉 공급망 관리(SCM) 모형 예시



자료 : Singh, A., et al. "Application of Big Data in Supply Chain Management", *Materials Today: Proceeding*, Vol. 4, No. 2, Part A, 2017, p.1107

⁵¹⁾ Singh, A., D. Jain, I. Mehta, J. Mitra and S. Agrawal, "Application of Big Data in Supply Chain Management", *Materials Today: Proceeding*, Vol. 4, No. 2, Part A, 2017

⁵²⁾ Witkowski, K, "Internet of Things, Big Data, Industry 4.0 - Innovative Solutions in Logistics and Supply Chains Management", *Procedia Engineering*, Vol. 182, 2017

Wang(2016)⁵³⁾은 빅데이터를 활용한 시장동향, 고객 구매패턴 및 유지·보수 주기 파악 등으로 비용절감과 합리적 비즈니스 의사결정 지원에 주목하였다. 인터넷을 통해 생산되고 전달되는 데이터의 양이 급격히 증가하고 있는 현대의 비즈니스 환경에서 BDBA(Big Data Business Analytics)의 중요성을 인식하고, 물류 및 공급망관리(LSCM) 분야에 BDBA를 적용한 선행 연구를 검토·분류하였다. 또한 기업의 관리자는 BDBA와 SCM을 기업의 전략적 자산으로 받아들이고 통합 엔터프라이즈 비즈니스 분석을 수행할 것을 강조하고 있다.

Nguyen(2017)⁵⁴⁾은 BDA가 SCM의 어떤 영역에 어떻게 적용되었는지를 파악하기 위해 Mayring(2008)이 제안한 내용분석법(content analysis method)을 이용하여 선행연구를 분류하였다. 분류 분석은 BDA가 적용된 SCM의 영역, BDA 분석 정도, 사용된 BDA 모형의 유형, 모형 개발에 사용된 BDA 분석방법론 등의 질문에 기반하고 있다.

Tsao(2017)⁵⁵⁾는 SCM하에서 공급자(supplier)와 소매업자(retailer)의 연결 고리에 주목하고, BDA를 활용함으로써 소비자(customer)의 채무불이행 위험을 완화시킬 수 있다고 주장하였다. 특히, 신용기간을 결정하는 공급자 또는 소매업자는 BDA의 효과가 일정 임계값 이상일 때 단독으로도 BDA를 적용하는 것을 선호하며, BDA 비용이 분담될 때 신용기간, 주문 수량과 BDA 효과가 더욱 증가하는 것으로 나타났다.

2) 금융 리스크 관리(Finance Risk Management)

금융회사는 리스크를 평가하고 그 결과를 바탕으로 의사결정을 수행한다. 이론적으로 더 많은 정보의 획득은 리스크 평가의 신뢰도를 높일 수 있다. 이 관

⁵³⁾ Wang, G., A. Gunasekaran, E. Ngai and T. Papadopoulos, "Big Data Analytics in Logistics and Supply Chain Management: Certain investigations for research and applications", *International Journal of Production Economics*, Vol. 176, 2016

⁵⁴⁾ Nguyen, T., L. Zhou, V. Spiegler, P. Leromonachou and Y. Lin, "Big Data Analytics in Supply Chain Management: A state-of-the-art literature review", *Computers and Operations Research*, In press, <https://doi.org/10.1016/j.cor.2017.07.004>, 2017

⁵⁵⁾ Tsao, Y., "Managing Default Risk under Trade Credit: Who should implement Big-Data analytics in supply chains?", *Transportation Research Part E: Logistics and Transportation Review*, Vol. 106, 2017

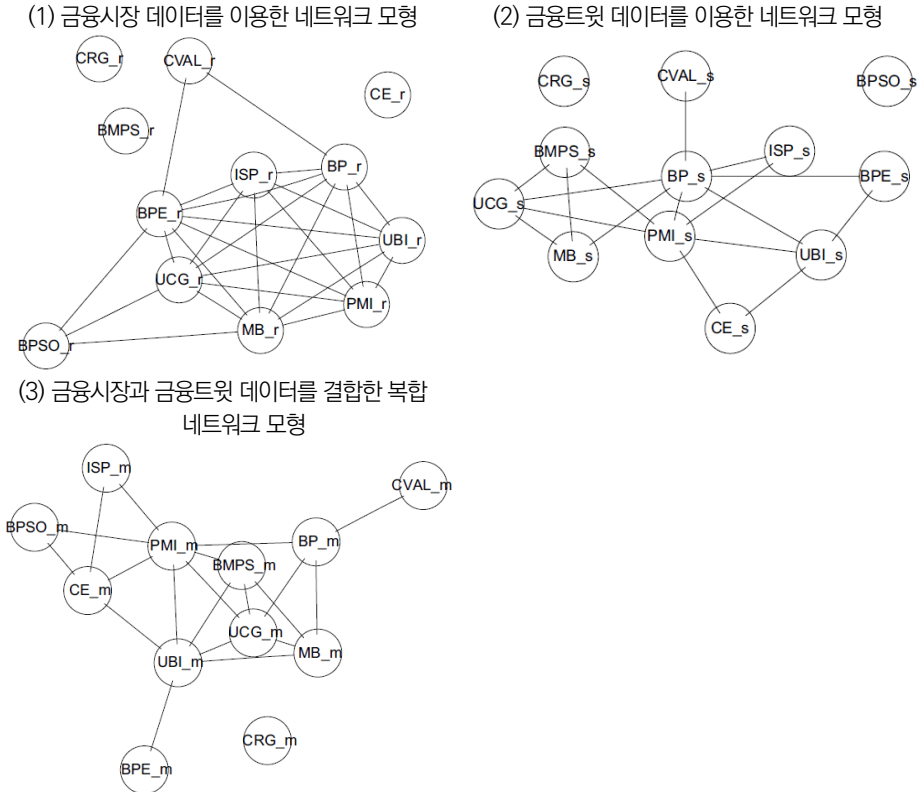
점에서 빅데이터 분석은 오늘날 금융회사의 의사결정자에게 필수이며, 향후 모든 금융 관련 의사결정은 빅데이터 분석에 바탕을 둘 것이라는 전망이 나온다. 이에 EIU(2014)⁵⁶⁾는 전 세계 208명의 금융회사 재무회계 책임자를 대상으로 조사를 실시하였다. 금융회계 분야에서 빅데이터가 가져올 가장 큰 변화로는 직관적으로 연관성이 보이지 않는 외부환경 변화를 실시간으로 해석할 수 있는 기능, 자본 비용 시나리오의 정확성 향상 및 회계 데이터와 보고서의 자동생성 등이 꼽혔다. 리스크 관리에서 빅데이터가 가장 성공적으로 활용된 분야로는 신용카드사기 방지, 대출 채무불이행 방지, 유동성 요건 충족 및 시장동향 예측 순으로 나타났다. 특히, 실시간 분석을 기반으로 의심스런 거래를 판단하여 고객에게 신속하게 연락할 수 있게 함으로써 신용카드사기를 방지하며, 행동 모니터링을 통해 채무불이행 위험을 예상함으로써 리스크에 선제 대응할 수 있게 하는 것으로 나타났다.

Cerchiello(2016)⁵⁷⁾는 기존 모형과 달리 금융시장 가격에 포함된 정보뿐 아니라 재무관련 트위터 데이터를 사용하는 체계적 리스크관리 모형(systematic risk model)을 개발하였다. 제안된 방법론은 금융 시장과 금융 트윗의 두 가지 데이터 소스를 베이지안(Bayesian) 접근법을 사용하여 결합하고 체계적 리스크관리 모형을 추정한다. 저자에 따르면 이는 빅데이터에 기초한 최초의 체계적 리스크관리 모형이며, 다음 <그림 3-15>와 같이 금융 기관들 간 상호 연관 관계를 보여준다.

⁵⁶⁾ Economist Intelligence Unit, "Retail banks and big data: Big data as the key to better risk management, The Economist, 2014

⁵⁷⁾ Cerchiello, P. and P. Giudici, "Big Data Analysis for Financial Risk Management", *Journal of Big Data*, <https://doi.org/10.1186/s40537-016-0053-4>, 2016

〈그림 3-15〉 이탈리아 은행의 네트워크 모형



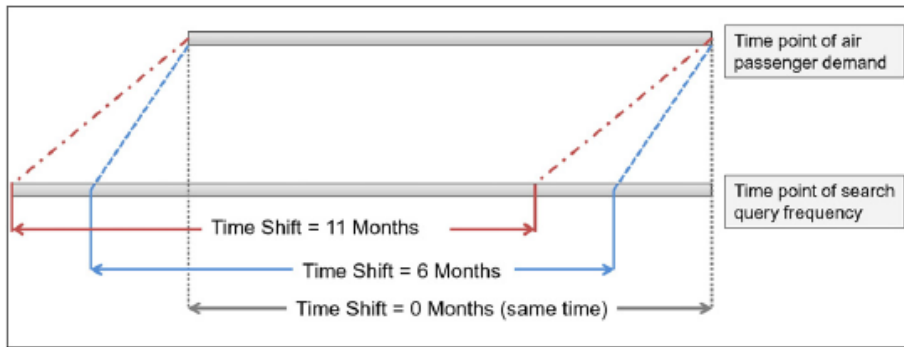
자료 : Cerchiello, P. and P. Giudici, "Big Data Analysis for Financial Risk Management", Journal of Big Data, <https://doi.org/10.1186/s40537-016-0053-4>, 2016, pp.9~11

3) 항공여객 수요 전망(Air Passenger Demand Forecast)

항공여객 수요예측은 공항의 운영계획 수립에 영향을 미치는 중요한 요소 중 하나이다. 공항 운영사는 장기수요예측을 통해 장기적인 운영계획을 수립하여야 하며, 동시에 단기수요예측을 통해 즉각적인 상황들에 대처하여야 한다. 하지만 대부분의 공항 운영사는 단기수요예측에 있어서 개별 전문가의 경험에 의지하고 있어 전문가의 숙련정도에 따라 예측 정확도의 변동성이 심하다. 이에 Kim(2016)⁵⁸⁾은 인터넷 검색어 기반의 빅데이터를 활용하여 전문성 수준과 관계없이 신뢰할 수 있는 단기수요예측 모형을 개발하였다. 제안된 예측모형은 네

이버 검색빈도 데이터를 사용하여 인천국제공항의 여객수요예측에 적용된 결과, 5.3%의 평균예측오차로 검색 빈도수가 0.1% 증가함에 따라 8개월 후 약 195,000명의 승객이 증가함을 보였다.

〈그림 3-16〉 예측모형의 시간이동(time shift) 개념도



자료 : Kim, S. and D. H. Shin, "Forecasting Short-term Air Passenger Demand using Big Data from Search Engine Queries", *Automation in Construction*, Vol. 70, 2016, p.102

⁵⁸⁾ Kim, S. and D. H. Shin, "Forecasting Short-term Air Passenger Demand using Big Data from Search Engine Queries", *Automation in Construction*, Vol. 70, 2016

제4장

해운빅데이터의 파악 및 확보 《

제1절 현황분석

1. 분석 방법

빅데이터의 소재를 원천별로 파악하기에 앞서 해운기업들이 빅데이터에 대해 어떻게 인지하고 있는지, 빅데이터를 실제 업무에 활용하고 있는지에 대하여 설문조사를 통하여 확인하였다.

설문은 전통적 데이터와 시장분석에 대한 설문의 추가적인 항목으로 진행되었으며 인지도에 대한 질문 두 개 항목과 활용에 대한 질문 한 개 항목으로 구성되었다. 설문지는 11개 회사의 68명에게 배포되었으며 설문대상은 해운기업의 주요 영업의사결정을 담당하는 자로서 다양한 조직계층에서 선정되었다. 개별응답자의 선정은 각 기업의 영업 또는 기획담당임원에 의해 이루어졌다. 설문조사의 구체적인 사항은 부록에 기술되어있다.

2. 현황

1) 빅데이터에 대한 인지도

빅데이터는 도입과 활용의 역사가 길지 않고 해운기업이 활용한 사례가 많지 않아 많은 사람들이 개념적인 이해도 못하고 있는 경우가 많다. 이를 확인하기 위하여 다음의 질문을 제시하고 해당 항목을 모두 선택하도록 하였다.

빅데이터와 전통적 데이터의 차이를 알고 있다.

빅데이터 분석의 기법들에 대하여 적어도 개념적인 이해를 하고 있다.

빅데이터 분석과 4차 산업혁명의 차이를 설명할 수 있다.

이에 대하여 전체 응답자의 약 1/2에 해당하는 응답자가 빅데이터와 전통적 데이터의 차이를 알고 있다고 답했고 56%가 빅데이터 분석의 기법들에 대한 개념적인 이해를 하고 있다고 답했다. 빅데이터 분석기법에 대한 이해가 첫 번째 문항의 응답보다 높은 것은 이례적인 결과로 생각된다. 데이터의 차이를 명확히 이해하지 못하는 응답자보다 빅데이터의 수집, 분산파일시스템 또는 분산처리기술 등 분석기법에 대한 개념적인 이해를 하고 있다고 응답한 응답자가 더 많다는 것은 질문을 이해하고 응답하는 과정에서 오류가 있었을 가능성을 내포하고 있다.

한편 빅데이터와 4차 산업혁명의 차이를 묻는 질문에는 15%에 해당하는 8명이 차이를 설명할 수 있다고 응답했다.

빅데이터가 경제·사회 전반적으로 유행처럼 다루어지고 있음에도 불구하고 전체적으로 해운기업에서 빅데이터에 대해서 인지하는 수준은 구체적이거나 정확하지 않은 것으로 판단된다.

2) 빅데이터 활용현황

해운업에 종사하는 응답자들이 빅데이터가 해운업에 미치는 영향에 대하여 어떻게 생각하는지를 확인하기 위하여 다음의 질문을 제시하고 해당되는 내용을 모두 선택하도록 하였다.

빅데이터가 해운업에 미치는 영향에 대하여 생각해 본 적이 있다.

빅데이터 분석이 해운 의사결정에 긍정적인 영향을 미칠 것으로 생각한다.

해운 의사결정에 있어서 빅데이터 분석은 거의 무의미하다고 생각한다.

첫 번째 항목에 대해서 응답자의 60%에 해당하는 33명이 빅데이터와 해운업을 연관시켜 생각해 본 적이 있다고 응답하였다. 중요한 것은 빅데이터가 경제 및 사회의 다양한 분야에서 광범위하게 다루어지는 상황 속에서도 10명 중 4명은 해운업과 빅데이터 분석의 관련성에 대해 생각해 보지 않았다는 것이다.

두 번째와 세 번째 항목은 상호 배타적인 질문이다. 빅데이터분석이 해운 의사결정에 미치는 영향에 대하여 71%에 해당하는 39명이 긍정적이라고 답했으며 무의미한 것이라고 답한 응답자는 두 명에 불과했다.

3) 빅데이터 활용의 내용

국내 해운기업에서 빅데이터를 활용하는 내용을 확인하고자 빅데이터의 수집, 분석, 의사결정에 활용 여부를 묻고 각각에 해당하는 항목을 기술하도록 요구하였다.

이에 대하여 빅데이터를 수집하고 있다고 응답한 응답자는 전체의 20%에 해당하는 11명이었고 빅데이터를 분석하고 있다고 응답한 응답자는 10명(18%), 빅데이터를 의사결정에 활용하고 있다고 응답한 응답자는 6명(11%)이었다.

하지만 수집하는 항목을 기술한 것을 보면 빅데이터의 원래의 정의에 해당하는 것은 전혀 없고 모두가 전통적 데이터의 범주에 드는 것이었다. 아래는 수집 항목으로 열거된 것들이다.

선박공급량, 폐선량
선박
화물
신조
선사별 물동량
선가, 운임, 운임지수
선대증감
해운관련 뉴스, 보고서
성약
시황전망 등

분석기법 또한 빅데이터 처리를 위한 분산저장, 처리기술은 언급되지 않았고 시계열 분석, 기술적 분석, 추세분석, 통계분석 등 전통적 방법이 열거되었다.

활용영역으로는 용선, 운임의 결정, 신규시장개척, 신규사업진출, 선복 확보 및 장기계약체결, 선박매매 등 통상적인 해운업 의사결정의 범주를 넘지 못하였다.

결과적으로 원래의 개념에 부합되는 빅데이터를 수집, 분석, 활용하는 기업은 국내에 거의 없는 것으로 판단된다.

제2절 원천별 해운 빅데이터의 파악

1. 해운 관련 기관

1) 정부의 추진현황

정부는 최근 ‘정부 3.0 추진 기본계획’을 발표하였고 60개 이상의 정보공유 및 시스템 연계·통합과제와 빅데이터 시범과제를 추진하였으며 과학기술 기본 계획에서 빅데이터를 30대 중점기술로 선정하기도 하였다. 미래창조과학부는 최근 빅데이터 활용을 통해 사회문제를 해결하고자 일자리, 청소년 복지, 소상공인지원, 보건의료 분야에서 빅데이터 사업화 컨설팅 과제를 지원하고 있다.⁵⁹⁾

2016년 10월 해양수산부는 해양수산 빅데이터 추진계획을 마련하였다. 이는 빅데이터가 새로운 패러다임으로 대두됨에 따라 예측에 근거해서 해양수산정책을 수립하고 관련 산업을 육성하기 위해 수립된 것이다. 이 계획에는 4개의 추진전략과 15개의 과제가 포함되었는데 그 항목은 다음과 같다.

⁵⁹⁾ 주해중·김혜선·김형로 (2017), p. 11.

〈표 4-1〉 해양수산 빅데이터 추진계획 개요

해양수산 빅데이터 민·관 공동활용을 위한 플랫폼 구축	플랫폼 구축
해양수산 빅데이터 인프라 구축	해양공간 활용모델
	연안 생태계 변동예측
	빅데이터 활성화 정책추진
단기성과 창출을 위한 성공사례 발굴	AIS를 활용한 항만운영효율 향상
	유해생물 출현 및 확산 예측
	어업지도 효율향상
	태풍피해 예측 및 대응
	위험예측 및 대응
중장기 과제추진으로 빅데이터 확산	수급 및 작황예측
	응용유전자 개발
	어항 예측도 향상
	항만물류 리드타임 단축
	선박의 최적 경제운항 모델개발
	항만시설 수급예측도 향상

자료 : 해양수산부, 해양수산 빅데이터 추진계획(안) 2016. 10, 저자 재정리.

해양수산부가 마련한 추진계획에는 해운업과 유관 산업의 의사결정의 질을 개선하기 위한 빅데이터의 활용에 대한 내용은 포함되지 않았다.

2) 정부 및 공공기관 데이터

해운시장분석에 활용될 수 있는 원천데이터(primary data)를 생성하는 대표적인 공공기관으로는 해양수산부/항만공사와 관세청을 들 수 있다.

해수부/항만공사는 항만운영정보시스템(Port-MIS)라는 형태로 전국 28개 무역항을 대상으로 선박입항신고, 항만 내 시설 사용, 관제사항, 화물 반출입, 세입징수, 출항신고 등과 관련된 자료를 처리하고 있다. 항만운영정보시스템은 28개 항만을 나누어 운영하고 있었으나 금년부터는 통합시스템으로 운영되고 있다.

이 중 아래자료는 공공개방데이터로서 오픈 API의 형태로 민간에 제공되고 있다.

해양수산통계연보-국가승인통계

컨테이너수송실적
선박입출항실적
화물처리실적
컨테이너반출입정보
항만시설 사용정보
선박운항정보
항만국통제정보
해양사고정보
선박통계

또한 해운항만물류정보센터 SP-IDC를 통해 제공되는 정보도 활용이 가능하다.⁶⁰⁾ 해운항만물류정보센터는 정책담당자와 물류주체들의 의사결정을 지원할 목적으로 설립되어 관련 자료를 제공하고 있으며 아래와 같은 구조로 시스템이 설계되어 있다.

〈그림 4-1〉 SP-IDC 시스템 개념도



자료 : www.spidc.go.kr

⁶⁰⁾ www.spidc.go.kr을 통해 정보가 제공된다.

제공되는 정보로는 Port-MIS를 기초로 한 해운항만통계, 등록선박정보, 국제물류 통계 등이 있다. 제공되는 정보의 구체적인 구조는 별책에 정리되어 있다.

별책자료에서 볼 수 있듯이 SP-IDC에서 제공하는 연간 또는 월간 데이터이며 대체로 90년대 초반부터의 시계열 자료형태로 제공된다.

국가물류통합정보센터⁶¹⁾를 통해 제공되는 정보는 원천데이터가 아닌 다른 기관(예를 들면 Port-MIS)의 자료를 취합·정리하여 제공하는 것이므로 자료원으로서의 의미는 크지 않은 것으로 볼 수 있다.

관세청 자료의 경우 관세청 홈페이지⁶²⁾와 공공데이터 포털을 통해 다음의 데이터를 시계열로 제공하고 있다. 시계열 데이터에는 수출입 금액과 중량이 포함되어 있다. 중량자료에는 TEU환산, 부피화물의 변환 등 별도의 처리가 필요하다는 한계가 있다.

수출입 총괄

품목별, 국가별, 성질별, 신 성질별, 품목별·국가별, 대륙별, 경제권별, 세관별, 종류별, 항구·공항별 수출입실적

데이터의 제공을 위해 관세청은 공공데이터 개방 책임관을 두고 있으며 실무담당자 2명이 데이터 제공업무를 담당하고 있다

공공데이터 포털을 통한 관세청 데이터의 제공은 개방형 API의 형태로 이루어지고 있으며 개방형 API의 활용에 대한 구체적인 가이드도 별도의 파일로 제공하고 있다. 또한 상시 개방되지 않는 데이터는 정부의 정보공개 포털(www.open.go.kr)에서 ‘공공데이터 제공신청’을 통해 추가로 확보할 수 있다.

이러한 제도는 2013년에 마련된 ‘공공데이터의 제공 및 이용활성화에 대한 법률’에 근거하여 운영되고 있다.

여기에서 중요하게 다루어져야 할 부분은 해운시장분석을 전제로 한 자료의 유용성이다. 해운시장은 글로벌 시장이기 때문에 해운시장의 분석에 사용되는 데이터는 특정지역이나 한 국가에 국한되어서는 안 된다. 특정지역이나 국가의 자료는 글로벌 시장의 역동성을 설명하는 입력변수로 활용되는데 한계가 있기 때문이다.

61) 국가물류통합정보센터, www.nlic.go.kr

62) 국세청, www.customs.go.kr

우리나라가 세계 10위권의 무역대국이기는 하나 우리나라 항만의 입출항 물량에 대한 정보나 세관의 수출입 무역 통계는 해운시장 전체를 분석하는 모델의 입력변수로 활용되기에는 부족한 것이다. 하지만 우리가 주요 무역국의 통계를 대부분 확보할 수 있다면 문제는 달라진다. 최근 미국을 비롯한 선진국들이 데이터의 개방성을 증가시키고 있는 추세이기 때문에 앞으로 우리나라뿐만이 아니라 다른 주요 무역국의 교역량 자료를 실시간으로 확보할 가능성이 커지고 있다. 이러한 상황에서는 데이터의 유용성이 확보될 수 있으나 현재로서는 제한적임이 감안되어야 한다.

2. 해운기업의 데이터

해운기업의 데이터는 기업의 IT시스템에서 거래의 기록으로 생성되는 데이터와 기업이 선박이나 장비에 센서를 부착하여 수집하는 정보로 대별된다.

기업의 IT시스템의 중심에는 이해관계자에게 정보를 제공하기 위한 회계시스템이 있고 대부분의 하위시스템은 이 회계/자금시스템을 통한 재무제표의 산출 및 자금 관리를 지원하는 형태로 구성되어 있다. 영업이나 관리시스템이 운영의 효율성 제고를 위해 회계와는 별도의 모듈을 운영하는 것도 일반적인 현상이나 회계정보이든 영업 및 관리정보이든 기업의 입장에서는 매우 민감한 정보이기 때문에 이를 외부에서 확보하여 분석에 활용하는 것은 현실적인 방안이 아니다. 하지만 기업이 선박이나 장비에 감지기(sensor)를 부착하여 확보하는 정보는 영업기밀이나 회계정보와는 무관한 매우 지엽적인 정보이므로 공유에 의한 가치의 증대만 인식된다면 외부에서 분석에 활용하는 것이 가능할 것이다.

기업이 선박이나 장비로부터 확보하는 정보는 대부분 물리적 상태의 변화 또는 공간적 이동에 대한 기록이다. 조사에 의하면 국내 A선사의 선박관리 조직이 선박에서 수집하는 정보는 다음의 항목을 포함하고 있다.

일자(date, time)

위치(position)

속도(speed)

부피(volume)
 압력(pressure)
 온도(temperature)
 각도(angle)
 전류, 전압(current, voltage)
 주파수(frequency)
 가동여부(activity)
 회전수(revolution)
 통과량(flow)
 수위(level)
 밀도(density)

A선사가 별도의 센서를 부착하지 않고 수집하는 항목은 약 1,700여개이다. 실제 선박에서 별도의 장비 없이 수집할 수 있는 정보의 종류는 수만 가지가 되는 것으로 알려져 있으며 이러한 데이터는 선박 운항 효율의 개선이나 선박건조 시 기기의 개발, 개선 등에는 활용도가 높은 것으로 알려져 있다. A선사가 확보하는 항목 리스트의 일부는 별책부록에 예시되어 있다.

3. 데이터 공급 업체

1) Clarkson Research Services

온라인 기반의 Shipping Intelligence Network이라는 해운조선 및 무역 데이터와 분석보고서를 제공하는 Clarkson Research Services는 Clarkson Group을 구성하는 회사 중 하나이다. Clarkson Group은 1852년에 설립되어 165년의 역사를 가지고 있는 통합 해운서비스 회사로서 전 세계 49개의 사무실에 약 1,400명의 인력을 보유하고 있다.

Clarkson Research Services는 영국과 상해를 기반으로 정규직 연구원만 75명을 두고 해운, 조선 및 교역에 관한 다양한 데이터를 수집, 분석하고 있다.

서비스 영역은 크게 해운과 교역, off-shore와 에너지, 선박가치평가로 나뉘는데 선박가치평가는 별도의 자회사를 통해 서비스되고 있다.

동사는 약 50년 이상 데이터를 발간해 왔으며 정보는 온라인과 인쇄물을 통해 제공된다. 주요고객은 해운기업, 금융기관, 정부, 보험회사, 조선 및기자재 회사 등이며 규격화된 서비스 이외에 고객의 요구에 따른 맞춤형 서비스도 제공하고 있다.

SIN에서 제공하는 시계열 자료를 웹사이트에 접속해서 건별로 분석했다. 총 항목 수는 4,700여 건으로 이에는 다음의 자료가 세분되어 있다(구체적인 항목의 리스트⁶³⁾는 별책부록 참고)

운임지수

선형별, 항로별, 기간별 운임

신조선가, 중고선가, 해체선가

선박발주 잔량

신조계약량

선박해체량

선박거래량

화물별 해상물동량

유가, 자원가격, 산업생산, 환율 등 주요 경제지표

연료유가

위의 항목에서 보듯이 Clarkson Research에서 제공하는 정보는 빅데이터의 범주에 속하는 데이터가 아니고 전통적인 시계열 데이터이다. 하지만 향후 분석 모델을 개발할 때 중요한 입력변수로서 투입될 수 있는 자료들이기 때문에 구체적인 파악의 가치가 큰 자료라고 할 수 있다.

데이터의 관측주기는 일간, 주간, 월간, 분기, 연간으로 구분되어 있으며 단기 변동성의 측정이 중요한 데이터는 관측 주기가 짧고 그렇지 않은 데이터는 주간 또는 월간으로 측정되어 있다. 관측주기보다 긴 통계치는 짧은 관측치의 기간평균으로 작성되어 있다.

관측시점이 가장 오래된 데이터는 1970년부터 측정이 되었으나 신 선형의 경우 선형개발시점 이후의 데이터가 존재한다. 모든 데이터가 연속성을 가지고 현

⁶³⁾ www.clarksons.com/services/research/, 검색일자 2017.12.1

재시점까지 기록된 것은 아니다. 신선형의 개발로 기존 선형의 진부화가 이루어진 경우 관측이 종료된 자료도 다수 있어 향후 입력변수화 할 때 선형의 보정을 통하여 연속성을 확보할 필요가 있다.

측정 단위의 경우 운임지수, 일당 수익력, 톤당 운임 등이 필요에 따라 적절히 사용되었으며 건조량이나 발주잔량 등은 척수와 dwt로 측정이 이원화되어 있다.

2) IHS Markit

IHS Markit그룹은 다양한 산업과 신기술 분야의 종합 컨설팅업체로 성장 중이며, 과거 IHS-Fairplay사, 미국 Journal of Commerce 등의 합병을 통해 성장하였다. 특화 분야로는 재무, 경제, 국가별 투자 위험 평가, 에너지, 화학, 자동차, 해운과 무역(maritime and trade), 신기술 등이다. 해운분야에서는 미국 해상 수출입 화물에 대한 100%의 정보, 86개 국가와 협조를 통해 항만의 수출입 무역액, 항만 입출항 정보를 구축하고 있다. 매 3분마다 인공위성이 제공하는 전 세계 대부분 선박(100G/T 이상)에 대한 추적 정보를 항만 입출항 정보, 무역정보 등과 연계하여 글로벌 빅데이터를 구축 중이다.

해운과 무역 분야에서는 글로벌 공급망(supply chain)에 대한 위험성 평가와 효율성 개선을 목표로 한 다양한 빅 데이터 분석과 정보를 제공하고 있다⁶⁴⁾. 구체적으로 Maritime Solutions, Global Trade Solutions, Maritime Portal, Global Trade Atlas, Piers US Import Export Data, World Trade Service, Maritime Intelligence Risk Suite, Market Intelligence Network, Maritime & Trade Fleet Capacity Forecast, Maritime & Trade Freight Rate Forecast 등의 서비스를 제공하고 있다.

Maritime Solutions는 Maritime Portal, Bespoke Data, Lloyds Register of Ships, Port & Terminal Guide, Market Intelligence Network, Maritime Intelligence Risk Suite 등으로 세분된다. Maritime Portal은 세계 100 G/T 이상 20만 척의 선박에 대해 600여 분야의 정보를 제공하는 Sea-web과 인공 위성 정보를 활용한 선박위치 추적 AIS Live가 포함된다.

⁶⁴⁾ www.ihs.com/info/0516/maritime-trade-solutions.html(검색일자 : 2017년 9월 22일)

〈표 4-2〉 IHS Markit의 서비스 목록

Maritime & Trade: Maritime Solutions	Maritime Portal Ports & Terminals Guide Bespoke Data Solutions Maritime Intelligence Risk Suite Lloyd's Register of Ships / List of Shipowners & Managers Market Intelligence Network (MINT)
Global Trade solutions	PIERS Global Trade Atlas (GTA) World Trade Service (WTS) Market Intelligence Network (MINT)
Maritime Portal	Protect your assets: Automatically monitor any ship slowing down or anchoring in your custom zones or ports Identify new business: Receive notice when a ship has changed its AIS status to moored or aground Improve operational efficiencies: Use one tool to learn critical details about ships coming into your port and to receive alerts when a ship changes its destination or estimated arrival Target your marketing activities: Leverage real-time notifications of ship registration or ownership changes to ensure your internal databases are accurate Make contextualized decisions: Get a holistic picture of the world fleet, the companies that manage them, the ports they call at, their movements and trading history
Global Trade Atlas – Commodity Trade Data	Build a global view of a commodity or group of commodities See the effect of new market entrants Identify emerging markets, trends and supply sources Uncover supply-demand imbalances and emerging opportunities Compare your market share to the competition
PIERS US Import Export Data	Increase operational efficiencies with decision-ready data Make smarter decisions based on greater data accuracy Capitalize on market opportunities by tracking supply-demand shifts Benchmark competitor shipments, trading partners, suppliers and strategies Generate sales leads by targeting companies based on their trade activities
World Trade Service – International Trade Data & Forecasts	Make timely, profitable business decisions Uncover risks and opportunities Recognize market trends and key catalysts Benchmark the competition Support economic development and infrastructure investments
Maritime Intelligence Risk Suite (MIRS)	Reduce maritime exposures Make informed underwriting decisions Carry out compliance checks on ships and their seven levels of ownership Comply with anti-money laundering laws Establish strong domain awareness

자료 : <https://www.ihs.com/products>(검색일자 : 2017년 9월 22일), 저자 재정리

IHS Markit이 제공하는 데이터의 구체적인 리스트가 파악되었으며 이는 별책부록에 포함되어 있다. IHS Markit의 자료들은 상당부분 횡단면자료로 이루어져 있어 2단계 연구에서 입력변수로 활용할 때 활용도가 크지 않을 것으로 판단된다. 하지만 AIS 자료의 경우 특정 공간별 시계열자료를 활용하고 교역에 대한 자료를 결합할 경우 매우 가치 있는 결과를 도출할 수 있을 것이다.

3) FleetMon (Jakota Cruise Systems GmbH)⁶⁵⁾

FleetMon은 운송지연문제의 해결을 위한 AIS자료의 제공을 위해 2007년 독일에 설립된 회사이다. 등록 선박은 46만 척 이상이며 전 세계 164개국에 32만 명의 등록 고객을 두고 있다.

자료는 위성 AIS, 지상 AIS(Terrestrial AIS) INMARSAT C와 자체 개발한 S1 트랜스폰더를 통하여 수집한다. 전 세계 5,125개 항구, 1,200개의 AIS 스테이션에서 나오는 정보를 수집하고 있는데 수집하는 분량은 초당 3,500건으로 하루에 약 3억 건의 데이터가 축적된다. 과거 3년간의 시계열 데이터가 활용 가능하며 고객들에게는 API연동서비스가 제공된다.

원시자료를 제공하는 것에 덧붙여 초보적인 데이터 분석도 진행하고 있어 선박의 이동거리, 항해소요시간, 지연가능성, 최적항로 등을 예측 또는 분석할 수 있다.

4) Platts cFlow⁶⁶⁾

cFlow의 AIS는 전 세계 1천 dwt 이상 선박 약 6만 척의 선박동정을 매 30분마다 한 번씩 관측하여 동정을 파악한다. 이를 기반으로 원유, LNG, LPG, 케미컬, 철광석, 석탄, 컨테이너 등 주요화물의 이동현황을 파악하는데 이는 전 세계를 약 12,000개로 구분한 권역 간 이동을 파악함으로써 이루어진다. 화물 이동뿐만 아니라 성약 등을 실시간으로 파악하고 예측함으로써 화물의 모니터링, 체선방지 등에 활용한다.

서비스는 3개의 등급으로 나뉘어 제공되며 고객맞춤형 서비스의 제공도 가능

⁶⁵⁾ KMI (2017), 『The Ocean』, Vol.08, p.87.

⁶⁶⁾ www.platts.com/products/cflow, 검색일: 2017.12.1

하다. 서비스등급에 따라 최대 4년간의 과거 데이터를 제공받아 분석에 활용할 수 있다.

5) Baltic Exchange⁶⁷⁾

볼틱 해운거래소는 영국 런던에 본부를 두고 있으며 1744년에 설립되어 250년 이상의 역사를 가진 독립적인 해운시장정보기관이다. 해운시장 분석에 있어 가장 중요한 것은 1985년 이후 지속적으로 산출하고 있는 운임지수이다. 장기간의 시계열 자료로서 일관성을 가지고 독립적인 패널에 의해 유지되었기 때문에 시장분석을 위한 객관적 입력변수로서 탁월한 위치를 점하고 있다.

Cape, Panamax, Supramax, Handy, dirty tanker, clean tanker 등 선종별로 운임지수가 산출되며 이는 또 각 항로별 지수로 세분되어 지역별 시장의 움직임을 파악할 수 있게 한다.

또한 Baltic Exchange Forward Assessments를 통하여 독립 패널 브로커들이 제출한 일간 FFA 평가액을 공표함으로써 FFA 정산의 기준을 제시하고 있다.

운임지수와 FFA 평가액 공히 향후 시장분석에 있어 무엇보다도 중요하고 신뢰성 있는 원천정보가 된다. 특히 선물의 가격은 미래 spot가격에 대한 정보를 담고 있어 가격발견기능(price discovery function)을 활용한 다양한 분석이 시도될 수 있어 선물가격의 시계열 자료는 특별한 의미가 있다.

4. 해운 빅데이터의 특징

1) 국내 원천데이터의 부족

해운시장분석에 활용될 수 있는 국내 원천데이터는 Port-MIS, 관세청의 수출입 통관실적 등에 국한되어 있고 그나마도 국내라는 지역적 제한성 때문에 글로벌 해운시장의 분석에 입력변수로 활용하는 데에는 한계를 지니고 있다.

선사에서 수집·생성하는 선박운항 관련 기술자료 등은 빅데이터의 성격을 가지고 있고 운항효율의 제고 등 효율성을 지니고 있으나 개별기업의 데이터에 국

⁶⁷⁾Baltic Exchange, www.balticexchange.com

한되는 경우 시장분석에 사용되는 데에는 한계가 있다. 이는 사물인터넷 기반의 데이터 확보의 역사가 길지 않아 축적된 데이터의 절대량이 부족한 것과 기업들이 자사의 자료를 경쟁력의 원천으로 인식하여 공유하지 않음으로써 공유에 의한 가치의 증대가 이루어지지 않는 것이 원인인 것으로 분석된다. 국내에 원천 데이터가 부족한 것은 해외 상업적 공급자에 대한 의존도를 증가시키는 원인이 된다. 상업적 공급자에 대한 의존도에 대해서는 아래 4)에서 다룬다.

2) 전통적 데이터 위주

해운업 관련 데이터 중 빅데이터의 범주에 들어가는 자료는 그리 많지 않다. 빅데이터는 간단히 정의할 때 5테라바이트 정도의 이동식 디스크에 데이터가 다 들어가지 않을 때라고 정의되고⁶⁸⁾ 보통 3V(Volume, Velocity, Variety)의 요건을 충족시키는 데이터로 정의되기도 한다.

설문조사를 통해서도 파악되었듯이 해운기업들이 해운시장분석에 활용하는 대부분의 데이터들은 이 범주에 들지 않는 전통적 데이터이다. 분석기법을 확장하여 빅데이터 처리와 기계학습 기반의 도구를 적용하는 것을 전제로 하고 향후 가용한 데이터를 확보해 나간다고 할 때 통행량의 측정을 위한 이미지 데이터, 광범위한 지역의 물동량 흐름을 파악하기 위한 대규모의 수치 정보 등은 빅데이터의 범주에 속할 것으로 생각되나 아직 시장분석에 이러한 자료가 활용되고 있지는 않다. 현재의 데이터로서 빅데이터의 범주에 속하는 자료는 전 세계 선박의 동정을 실시간으로 파악하는 AIS(Automatic Identification System)데이터가 대표적이며 각 선사에서 선박으로부터 사물인터넷(IoT)을 활용하여 수집하는 선박운항 모니터링 자료는 전 세계 선박으로 범위를 확장한다면 빅데이터의 범주에 속할 것이다.

3) 데이터의 중복성

해운시장 분석을 위한 자료들은 상업적 기관에서 수집하여 판매하는 자료들이 대부분이다. 이러한 상업적 기관이 다수 존재하면서 중복의 문제가 생기게

⁶⁸⁾ KMI (2017), 『The Ocean』, Vol.08, p.87.

된다.

전통적인 데이터인 시계열 자료들도 동일한 데이터를 여러 기관에서 판매하는 경우가 많다. 기관에 따라 다소 관측치에 차이가 있지만 데이터의 일관성만 유지된다면 시장분석 목적으로 활용될 경우 문제가 되는 수준은 아니다.

AIS데이터의 경우 IHS Markit, FleetMon, Vessel Value, Platts(S&P Global) 등 많은 기업이 데이터를 인공위성과 항만의 스테이션으로부터 수집하여 판매하는 것으로 파악되었다. 이러한 중복성이 활용의 측면에서 어떤 문제를 야기하는 것은 아니며 오히려 데이터 소비자 입장에서는 경쟁에 의한 가격의 하락이라는 긍정적인 요소로 작용할 수 있는 특성이다.

4) 상업적 공급자 주도

전통적 데이터이든 빅데이터이든 데이터 공급을 해외의 상업적 공급자가 주도하고 있다. 전통적 데이터는 Clarkson Research와 IHS Markit이 주로 담당하는데 설문조사에서 파악되었듯이 국내 기업들은 Clarkson Research의 데이터와 보고서에 대한 의존도가 크다.

AIS 데이터의 경우 다수의 공급자들이 시장을 선점하기 위해 각축전을 벌이고 있는 상황이며 대부분의 공급자들은 단순히 데이터만을 수집하여 판매하지 않고 분석서비스를 부가하여 가치를 증대하고자 하고 있다. 예를 들어 IHS Markit은 전통적 시계열 모델과 인공지능 기반의 모델을 사용하여 시황을 예측하고 그 결과를 고객에게 제공하는 상품으로 시장을 공략하고 있다.

5) 공유 및 개방 미흡

국내 원천데이터의 부족을 다루는 부분에서 이미 언급되었듯이 국내 해운기업들은 사물인터넷을 통해 확보한 선박 관련 기술적 데이터들을 공유하지 않는 것으로 파악되었다. 이는 이런 유형의 데이터들을 운영의 효율성이라는 면에서 차별화할 수 있는 경쟁우위요소로 파악하기 때문이다. 많은 기업들이 동등한 가치의 데이터를 내어 놓고, 공유를 통한 가치의 증대에 대한 공통된 인식을 하고 있다면 공유가 활성화 되겠지만 현실은 그렇지 못하기 때문에 기업들이 아직은 공유보다는 독점에 의한 효과가 크다고 판단하는 것이다. 향후 전제조건들이 충

족되고 기업들이 적극적으로 공유함으로써 윈-윈 하는 상황을 맞게 될 것인지에 대해서는 아직 불확실하나 촉매역할을 할 수 있는 기관이나 조정자 역할을 하는 정부의 움직임에 따라 상황이 달라질 수 있을 것이다.

또 하나의 문제는 데이터의 개방성이다. 주요 선진국을 중심으로 공공데이터의 '공공성'에 대한 인식이 커지고 있고 이러한 인식을 바탕으로 공공데이터를 개방하는 방향으로 변화가 진행되고 있다. 우리나라의 경우 데이터의 개방정도가 미흡한 것이 현실이며 이를 개선하기 위해서는 제도적 법률적 기반의 정비가 선행되어야 할 것이다.

제3절 해운 빅데이터의 활용사례

1. 해운기업 사례

1) 현대상선⁶⁹⁾

현대상선은 2013년 IBM의 Netezza Data Warehouse Appliance를 도입해 각종 해사 정보와 항해 관련 데이터의 실시간 처리에 사용해오고 있다. 현대상선에서는 선박과 주고 받는 데이터 기준으로 전체 선대에서 매일 약 2GB의 데이터가 발생하고 있다. Netezza 도입 후 사용자 조회 속도와 데이터 가공 속도가 향상되었으며, 몇 시간씩 걸리던 데이터 가공 시간이 분 단위로 단축되었고 하루 전 데이터를 받아서 영업해야 했던 과거와 달리 준 실시간으로 가공된 데이터를 사용할 수 있게 됐다. 이에 따라 실시간에 가까운 데이터 가공으로 더 정밀해진 분석 결과를 얻고 있다. 기존에는 단순한 영업 실적 데이터도 두 시간 반 이상의 시간을 들여 분석해야 했는데 Netezza 도입 후 이를 30분 이내로 단축할 수 있게 됐다. 뿐만 아니라 하루 전 데이터가 아닌 분 단위의 데이터를 즉시 사용할 수 있게 되면서 데이터 분석 정확도가 이전보다 훨씬 높아졌다. 이

⁶⁹⁾ 강미주 (2015), 특집기획 해운분야 ICT 혁명, 새로운 해운시대 열리나 (2015.5.4.)

<http://www.monthlymaritimekorea.com/news/articleView.html?idxno=16027>의 내용을 바탕으로

저자 정리, 검색일 : 2017. 12. 05

런 도입 효과는 기존에 낭비 되었던 인프라를 재활용하는 결과로 이어졌다. 이전에는 분석을 하더라도 결과는 일반적인 수준의 의사결정에 그쳤지만 이제는 데이터 가공 속도가 단축되어 매우 정밀한 데이터 분석까지 가능해졌다는 것이다. Netezza 도입과 관련 현대상선 관계자는 “대량의 데이터를 빠른 속도로 처리하며 준 실시간 분석 데이터를 적시에 제공해 의사결정을 지원하도록 개선해 운영 중”이라고 밝혔다.

2) 폴라리스쉬핑⁷⁰⁾

폴라리스쉬핑은 2013년 6월 운항 부문에서 오퍼스 벌크 솔루션을 도입했으며 채산성 검토, 운항 수지 관리, 항차 분석 및 월 결산 업무 등의 작업을 빠르고 쉽게 처리할 수 있게 됐다. 또한 선박과 ‘오퍼스 벌크’와의 운항 보고서 연동을 통해 운항 속도, 연료 및 윤활유 소모량, 운항 비용 등을 다양하게 분석할 수 있게 됨으로써 운항의 효율성을 높이고 있다. 이는 사물인터넷 기반의 선박데이터 수집을 통해 이루어진 것이다.

3) 머스크(Maersk)⁷¹⁾

덴마크 선사 머스크는 해운업의 차세대 기술적 진보를 ‘빅데이터’로 보고 적극적인 활용을 추진 중이며 변화하는 ICT 혁명에 가장 빠르게 대응하는 모습을 보이고 있다. 머스크 측은 “빅데이터는 머스크 선대 성장을 위한 커다란 기회”라면서 “어떻게 데이터를 수집하고 새롭게 이해할 것인지와 선박의 기능을 예측하고 시스템의 비효율적인 핸들링을 피할 수 있는지 등을 연구하고 있지만 전체 빅데이터의 잠재력에 비하면 이는 매우 작은 부분”이라고 설명했다.

머스크는 MMT라는 기술팀에서 데이터 분석을 통해 선대 효율화 및 에너지 효율화, 운영비 최소화를 추구하고 있으며 동시에 신규 사업의 발굴을 추진 중

⁷⁰⁾ 강미주 (2015), 특집기획 해운분야 ICT 혁명, 새로운 해운시대 열리나 (2015.5.4.), <http://www.monthlymaritimekorea.com/news/articleView.html?idxno=16027>의 내용을 바탕으로 저자정리, 검색일 : 2017. 12. 05

⁷¹⁾ 강미주 (2015), 특집기획 해운분야 ICT 혁명, 새로운 해운시대 열리나 (2015.5.4.), <http://www.monthlymaritimekorea.com/news/articleView.html?idxno=16027>의 내용을 바탕으로 저자 정리 검색일 : 2017. 12. 05

이다. 또한 최신 기술을 활용하여 선박의 구상선수 및 프로펠러 등을 개조하거나 엔진을 미세 조정하여 선박의 효율성을 최대 7~8%까지 증가시키고 있다.

유지 보수에 있어서도 빅데이터의 실시간 활용은 더 나은 결정을 하게 해준다. 그 예로 음향 신호를 사용해 부식을 측정하고 다른 파라미터를 적용함으로써 장비의 업그레이드 및 교체 시점에 대해 더 정확하게 파악할 수 있다.

2. 물류기업 사례

1) 한화토탈⁷²⁾

한화토탈은 스마트폰 등 모바일기기에서 주요 제품의 주문 및 조회가 가능한 ‘스마트 오더’와 실시간 배송 추적을 지원하는 ‘스마트 물류(Smart Logistics)’ 시스템을 오픈했다. 제품별 주문 유형에 알맞게 주문 화면을 세분화 했고, 일부 제품은 재고와 여신상 문제가 없으면 주문과 동시에 출하까지 자동으로 진행될 수 있도록 했다.

종전 고객들은 데스크톱 컴퓨터에서만 주문·출하 업무를 할 수 있었으나 이번 모바일 시스템 오픈으로 고객들은 언제, 어디서든지 주문을 할 수 있게 됐다. 또한 SMS를 통해 주문된 제품의 배송 진행 정보를 실시간으로 제공받을 수 있으며, 구매 내역, 시험 성적서 등 정보도 바로 확인할 수 있다.

운송 기사는 모바일 어플리케이션을 통해 운송 할당 물량을 확인하고 한화토탈의 대산공장 출입 시 무인계근시스템으로 계근한 뒤, 공장 내 출하장까지 경로를 내비게이션으로 안내받을 수 있다. 운송 기사가 각종 출하 서류를 직접 발행할 수 있는 전용 키오스크도 설치해 출하 소요 시간을 대폭 줄였고, 공장을 나선 차량의 운송 경로와 상태는 GPS를 통해 실시간 모니터링 때 보다 안전한 배송을 지원한다.

한화토탈은 이외에도 고객에게 순정 제품을 정확하게 공급하기 위해 제품의

72) 조지민 (2017), 한화토탈, 스마트 영업·물류 시스템 오픈...모바일로 실시간 처리·조회 (2017.11.9.), <http://www.fnnews.com/news/201711090833201676>의 내용을 바탕으로 저자 정리, 검색일 : 2017. 12. 05

정품, 정량을 보증하는 전자 봉인 시스템을 적용한 운송 차량도 시범적으로 도입했다.

2) UPS (United Parcel Service)⁷³⁾

미국에 본사를 둔 세계적 물류 운송업체 UPS는 ORION(On-Road Integrated Optimization and Navigation)을 활용한 빅데이터 분석을 통해 최적의 배달 경로를 설정한다. ORION은 2억5천만 개의 주소 데이터를 매일 분석하고, 교통 상황, 날씨, 물동량, 고객 분포 등을 고려하여 분당 3만 개의 경로를 최적화한다. 경로설정 시 좌회전 대신 P턴을 하는 방식으로 최적화되어 배송 시간을 단축하고, 연료비도 절감한다. 좌회전을 하기 위해 신호를 기다리는 것은 연료와 시간을 낭비하는 것일 뿐만 아니라 사고 발생 가능성도 높인다는 것이 UPS의 설명이다.

배송 예측의 정확도가 높아지면서 고객들은 실시간으로 배송 시각이나 배송지를 변경할 수 있게 되었다. My Choice라는 모바일 앱을 통해 고객들의 변경요청을 받으면 기존의 경로는 자동으로 조정된다.

또한 UPS 차량의 엔진, 브레이크 등 주요 장치에는 센서가 부착되어 있는데 이 센서가 실시간으로 차량 정보를 서버에 전달한다. 전달된 정보를 바탕으로 이상 징후를 발견하면 차량이 고장 나기 전에 이를 점검하고 수리하여, 사고를 예방한다.

3) DHL⁷⁴⁾

DHL은 2009년 빅데이터를 도입해 매일 기록되는 배송 도착지, 크기, 무게, 내용물 등 수백만 건의 배송 정보를 통해 소비자의 물류 서비스 이용 흐름을 파

⁷³⁾ 임예리 (2017), 자동 라우팅, 물류효율 높일 밑거름 될까 (2017.7.16.), <http://logiseconomy.tistory.com/3322>, 검색일 : 2017. 12. 05

Choi, C. (2014), UPS, Big Data로 배송 적시성 확보하기 (2014.10.15.), <https://jacesky1.wordpress.com/2014/10/15/ups-big-data로-배송-적시성-확보하기/>, 검색일 : 2017. 12. 05

⁷⁴⁾ 정유진 (2014), 글로벌 물류기업, 빅데이터 적극 활용 (2014.3.9.), http://www.dt.co.kr/contents.html?article_no=2014031002011132748001의 내용을 바탕으로 저자 정리, 검색일 : 2017. 12.05

악하고 패턴을 분석하는 데 활용하고 있다.

빅데이터를 활용한 성공 사례로 회사는 스마트 트럭(SmartTruck)을 꼽는다. 스마트 트럭은 실시간 교통 상황, 수신자의 상황, 지리적·환경적 요소를 고려해 최적화된 배송경로를 실시간으로 제공하는 배송 차량이다. 스마트 트럭을 도입한 후, 가장 효율적인 경로를 분석해 배송을 진행하고 있으며 배송 실패율을 제로에 가깝게 만들고 있고 불필요한 연비를 줄일 수 있다고 회사 측은 설명했다.

또한 빅데이터를 장·단기 투자규모 예측에도 활용하고 있다. 회사 측은 "수요 증가를 예측해 허브 시설 확충, 물류 센터 확장, 차량 증편에 대한 투자규모를 결정하기 때문에 투자위험 부담을 최소화 할 수 있다"며 "빅데이터 분석으로 불필요한 지출을 억제하고, 하루 약 7,000명의 업무 성과에 해당하는 생산성을 낼 수 있을 것"이라고 말했다.

4) TNT⁷⁵⁾

TNT는 2013년 빅데이터에 이은 '테라데이터(Teradata)' 시스템을 구축하여 각 부서에 맞는 빅데이터 시스템을 활용하고 있다.

물류 프로세스를 담당하는 부서는 빅데이터를 분석해 배송시간을 단축, 비용 절감 및 정시 배송 서비스 개선에 활용한다. 예를 들어 각 단계별 소요시간, 정시에 배송 되는 비율, 각 국가별 처리 화물량 등 다양한 데이터가 배송서비스 개선에 적극 활용되고 있다. 고객 서비스 센터의 경우, 일정기간 동안 미발송된 화물에 관해 축적된 데이터를 기반으로 배송 관련 사고 또는 배송 지연 등을 미연에 예방한다. 마케팅 부서는 소비자의 이탈을 방지하고 소비자 지향적 서비스와 솔루션 상품을 개발하고 있다.

5) Amazon⁷⁶⁾

빅데이터 분석을 이용한 고객 이해와 구매 추천의 선구자인 아마존은 '예측

⁷⁵⁾ 정유진 (2014), 글로벌 물류기업, 빅데이터 적극 활용 (2014.3.9.), http://www.dt.co.kr/contents.html?article_no=2014031002011132748001의 내용을 바탕으로 저자 정리, 검색일 : 2017.12.05

⁷⁶⁾ 최재경 (2016), 빅데이터 분석의 국내외 활용 현황과 시사점, http://www.kistep.re.kr/c3/sub3_2.jsp?, 검색일 : 2017.12.05

배송'이라는 또 다른 파격적 행보를 시도할 계획인데, 이를 위해 2013년 12월 고객이 구매하기 전에 배송을 준비하는 '예측 배송(anticipatory shipping)' 서비스에 대한 특허를 취득한 바 있다.

'예측 배송'은 고객이 구매할지 여부가 불확실한 상황에서 고객 주소지 근처의 물류창고로 배송을 시작하는 것으로, 이는 기존 주문과 검색 내역, 위시 리스트와 쇼핑 카트에 담아놓은 상품, 반품 내역, 마우스 커서가 머무른 시간 등을 활용해 고객 자신보다 고객을 더 잘 이해하는 것으로 알려져 있는 아마존의 빅데이터 분석 역량에 바탕을 두고 있다. 이와 관련해 월스트리트저널(WSJ)은 '예측 배송은 방대한 고객 데이터를 제대로 활용하고 있는 아마존만이 가능한 서비스로 경쟁 업체들과 비교해 강력한 차별화 포인트가 될 수 있을 것이다'고 전망했다.

3. 항만 및 터미널 사례

1) 부산항만공사⁷⁷⁾

부산 항만 공사가 4차 산업혁명 시대에 대응하는 해운·항만·물류 데이터베이스와 분석 시스템 1단계 구축을 완료했다. 체계적으로 데이터를 축적·분석함으로써 선사들의 동향을 신속하게 파악하고 새로운 물동량을 창출하는 한편 물류 흐름을 방해하는 요인들을 찾아 해소하는 등 경쟁력을 높이는 게 목적이다.

빅데이터를 운영할 통합 센터에서는 항만 공사 각 부서에 흩어진 내부 데이터는 물론이고 부산항에 기항하는 국내외 선사들, 터미널 운영사들 물동량, 하역 시간, 선적 사용률, 트레일러가 터미널에 들어가 화물을 싣고 나오는 데 걸리는 시간, 컨테이너 장치장 현황 등 각종 정보를 모아 분석한다.

1차로 구축한 시스템에서는 화주와 운송대행업체(포워더)들이 원하는 선박을 손쉽게 검색해 선적 예약을 할 수 있는 기능을 갖췄다. 또, 수출입·환적 화물의

⁷⁷⁾ 이영희 (2017), 부산항만공사 해운물류 빅데이터 분석 시스템 구축, <http://www.yonhapnews.co.kr/bulletin/2017/09/14/0200000000AKR20170914152500051.HTML?input=1195m>, 검색일 : 2017.12.05

흐름과 변화, 특정 국가나 항로의 물동량 변화 등 이상 징후, 물동량 장단기 예측, 항로별 운임 현황 등도 파악할 수 있다. 인터넷은 물론 모바일 기기에서도 이용할 수 있어 이용자들의 편의를 높였다.

2019년 2, 3차까지 단계 구축을 마치면 부산항의 모든 물류를 언제 어디서나 한눈에 파악하고 변화에 신속하게 대응함으로써 경쟁력을 한층 높게 될 것이다. 부산항에 기항하는 국내외 선사들의 선박 운항 정보를 한곳에서 검색할 수 있어 화주나 운송대행업체들이 각 선사의 홈페이지를 일일이 찾지 않아도 쉽게 원하는 선박을 찾아 예약까지 가능하도록 하는 기능도 갖출 방침이다. 이렇게 구축한 데이터베이스 중 선사와 터미널 운영사 등의 영업 비밀에 속하지 않는 부분은 관련 기업들이 활용해 시너지 효과를 낼 수 있도록 할 전망이다.

2) 한진 부산 컨테이너 터미널⁷⁸⁾

한진부산컨테이너터미널(HJNC)은 유비쿼터스를 실현한 최첨단 자동화 터미널로 첨단 IT 인프라 네트워크 체계로 신속한 작업 환경을 갖추고, 무인 자동화와 전사적 작업 지시를 아우르며 생산성 극대화에 초점을 맞췄다. 특히 게이트 운영의 경우 RFID 차량 인식 기술과 영상 촬영 시스템을 이용하여 원활한 반출입 작업을 수행하고 있다. RFID 또는 바코드를 인식하여 차량 정보를 확인하고, 영상 촬영으로 컨테이너 외관 손상 여부를 검사한다. 사전정보를 통한 정확한 컨테이너의 반출입 승인뿐 아니라 스마트폰을 이용한 컨테이너 위치 정보 및 사전 정보 확인으로 대기 시간을 최소화하고 있다.

야드작업의 경우 게이트 반출입 또는 선박의 양하 및 선적되는 컨테이너의 효율적인 장치 능력을 확보하고 있다. 내외부 차량의 블록 진입시 RFID에 인식됨에 따라 크레인 사전 준비를 유도하고, 지속적인 크레인 작업으로 인해 작업 처리 속도 및 소비 시간을 확보할 수 있다. 시스템 관리 체계에 따른 '오토 리핸들링'으로 불필요한 이적 작업을 최소화하고, 운영 시뮬레이션으로 현 야드 상황을 토대로 미래 시점의 비효율적인 작업 발생을 사전 방지하고 있다. 본선 운영

⁷⁸⁾ 강미주 (2015), 특집기획 해운분야 ICT 혁명, 새로운 해운시대 열리나 (2015.5.4.),

<http://www.monthlymaritimekorea.com/news/articleView.html?idxno=16027>의 내용을 바탕으로

저자 정리, 검색일 : 2017. 12. 05

도 효율적인 선석 운영과 최적화된 하역 작업 계획으로 생산성을 제고하고 있다. 무선 단말 장치를 통한 작업 지시로 야드 트럭의 신속한 이동을 유도하고 본선과 야드 트럭간의 최단 거리 계산으로 이동 거리를 축소하고 있다.

3) 싱가포르 해운항만청⁷⁹⁾

‘해상 이벤트 인식을 위한 센스 기반 분석’의 약어인 SAFER(Sense-making Analytics For maritime Event Recognition)는 싱가포르 해운 항만청(MPA)과 IBM사가 협업으로 싱가포르항의 증가하는 선박 통행량을 관리할 목적으로 새로운 분석기반 기술을 접목시킨 시스템을 개발하고 테스트하고 있는 프로젝트이다. 이 SAFER 프로젝트는 이전에는 인적 관찰, 보고, 초단파(VHF) 통신, 데이터 입력에 의존했던 주요 작업의 정확성을 증가시키고 자동화를 위한 새로운 기능을 제공한다.

SAFER에는 총 7가지 모듈이 있는데, 이번 시험운영은 자동선박운행감시, 위반행위분석, 도선사탑승감지의 3가지 모듈을 대상으로 했다. 2017년 9월부터 점진적으로 시행할 SAFER 프로젝트의 나머지 모듈 시험은 2018년 1월에 끝날 예정이다.

첫 번째 모듈은 자동선박운행감시(Automated Vessel Movement Detection)이다. 선박 이동의 감지와 예측을 위한 인식과 분석기술을 이용하여, 새로운 모듈은 MPA와 선장간의 무선통신을 줄이고 실시간으로 선박의 위치와 입출항 시간을 자동 감지함으로써 선박 이동 정보를 입력할 필요가 없도록 한다. 부가적으로 SAFER는 이동시간, 위치 정보의 정확성을 34% 향상시켜 직원들이 다른 역할을 수행할 수 있도록 한다.

두 번째 모듈은 위반행위분석(Infringement Analytics)이다. SAFER 시스템을 위해 개발된 기계학습 기반 분석과 선박 예측 모델은 알람을 통해 의심스럽거나 비정상적인 선박 행위를 감지할 수 있어 업무효율성을 개선할 수 있도록 한다.

⁷⁹⁾ 이연경 (2017), 싱가포르항-IBM사, AI기반 해운·항만운영 개선 시스템 시험운영 (2017.9.7), <http://www.kmi.re.kr/web/trebook/list.do?rbsIdx=135>의 내용을 바탕으로 저자 정리, 검색일 : 2017.12.05

세 번째 모듈은 도선사 탑승 감시(Pilot Boarding Detection)이다. SAFER 모듈은 MPA가 자동적으로 도선사 탑승 시간을 감지하여, PSAM의 도선 서비스 수준을 검증할 수 있도록 한다.

2008년 1월에는 SAFER의 나머지 4가지 모듈도 가동되었다. 4가지 모듈은 싱가포르항 내의 교통밀도에 관한 고급정보(Utilisation Prediction), 불법적인 병커링 감시(Illegal Bunkering Detection), 허가되지 않은 지역으로의 선박 출입 감시(Detection of Vessel Entering Prohibited Areas), 선박 도착시간 예측(Prediction of Vessel Arrival Time)에 관한 것이다. SAFER의 7가지 모듈은 초대형 선박 서비스를 제공할 TUAS 컨테이너항에서 운영되고 있다.

4) Inchcape Shipping Services(ISS)⁸⁰⁾

해양 서비스 제공 업체인 Inchcape Shipping Services(ISS)의 전임 CEO Claus Hyldager는 데이터 기반 해양 서비스 회사인 Nisomar를 출범시켜 광범위한 비즈니스에 분석 기능을 제공하고 있다. Nisomar의 CEO인 Hyldager는 선박 소유자, 운영자, 중개인 및 용선자 및 금융 기관이 내리는 의사 결정을 지원하기 위해 빅데이터를 이용함에 따라 회사에 도움을 줄 수 있을 것이라고 말했다.

Nisomar의 서비스는 입항이나 출항 직전에 통제보다는 항해를 다루는 주요 데이터를 제공함으로써 항만관련 조직(기관)의 효율성 증대를 목표로 하고 있다. 이를 위해 고도로 숙련된 해양, 공급망 관리 및 기술 전문가 팀을 구성했다.

신기술은 해상 부문에서 중요한 선박 데이터, 무역 흐름 및 화물 활용, 항만 혼잡, 선박 라인업 및 처리 효율성 등의 다양한 중요한 데이터에 활용(접근)할 수 있다. Nisomar는 혁신적인 데이터 분석과 통찰력이 결합했을 때, 선주와 선박관리자, 운영사에게 전략적 컨설팅, 부문별 물류관리서비스를 통해 성공을 보장할 수 있음을 알게 되었기 때문에 설립하였다.

⁸⁰⁾ Port Technology (2017), Start Up Aims Big Data at Shipping Sector (2017.4.24.)

https://www.porttechnology.org/news/start_up_aims_big_data_at_shipping의 내용을 바탕으로 저자 정리, 검색일 : 2017.12.05

4. 시사점

4차 산업혁명만은 경제전반에 걸쳐 광범위하게 진행되고 있다. 사물인터넷(IoT), 빅데이터, 인공지능을 활용한 분석 등 최신의 기법들을 기업들은 앞다투어 도입하여 경쟁력을 제고하는데 적극적으로 활용하고 있다.

특히 물류기업은 최적화를 실현하기 위한 기반으로 빅데이터를 수집하여 분석하고 있으며 그 결과로 서비스 차별화를 도모하고 있다.

싱가포르와 같이 정부의 주도권이 강한 국가에서는 정부차원에서 산업의 인프라 구축 차원의 빅데이터 분석이 추진되고 있음을 알 수 있다. 사례에 소개되지는 않았지만 중국의 경우에도 지방정부와 학계가 결합하여 공동연구를 추진하고 있는 것으로 알려져 있다. 해운에서 빅데이터와 인공지능기반의 분석이 추진되는 사례는 그리 많지 않다. IHS Markit에서 일본선사들의 요구로 시황예측 시스템에 AI기반의 분석을 추가하여 개발을 완료했으며 최근에 업계를 대상으로 마케팅을 시작한 것으로 알려져 있다. 중국의 상해국제해운연구소(SISI : Shanghai International Shipping Institute)도 AIS 등 빅데이터를 활용하여 해운시황을 예측하는 연구를 진행하고 있다.

해운영역에서는 선박데이터를 활용한 운항효율성의 개선분야에서 진척을 보이고 있으며 물류영역에서는 빅데이터 분석을 통한 물류시스템의 최적화가 시도되고 있다.

이러한 상황에서 해운의 핵심의사결정인 거래의사결정을 지원하는 시스템에 대한 연구를 진행하는 것은 시의적절한 것으로 판단된다.

시장분석에 있어서 빅데이터의 활용성에 한계가 있는 것이 현실이므로 데이터 분석의 범주를 비단 빅데이터 뿐만아니라 전통적 데이터의 영역까지 포함하여 진행하는 것이 바람직할 것이다. 따라서 이 일련의 연구는 사례에 소개된 '빅데이터'분석에 초점을 맞추기 보다는 빅데이터와 전통적 데이터를 포괄하는 입력변수들을 선정하고 이를 AI기반의 분석도구를 활용하여 분석함으로써 예측이나 판단의 실질적 성과를 제고하는 방향으로 진행될 것이다.

제4절 빅데이터의 확보방안

빅데이터이든 전통적 데이터이든 데이터를 확보하는 방법은 다양한 형태로 존재한다. 각각의 방법에 대하여 실제 구현가능성이나 유효성이라는 관점에서 살펴볼 필요가 있다.

데이터의 확보방법으로 제시될 수 있는 대안으로 다음과 같은 방법을 검토한다.

사물인터넷 기반의 수집

데이터 풀(pool)의 형성

개방형 데이터에 연결

수탁이나 공동연구를 통한 확보

외부데이터의 구매

웹 탐색

이해당사자 간의 교환

1. 사물인터넷 기반의 수집

해운기업이나 연구기관이 사물인터넷을 통해 원천데이터를 확보하는 방법에는 두 가지가 있다. 하나는 해운기업이 운영하는 선박이나 장비에 센서를 부착하여 데이터를 확보한 후 인터넷망을 통해 주기적으로 전송받는 방식이고 다른 하나는 도로나 터미널 등 물류의 주요 간선망이나 거점에 센서를 부착하여 데이터를 수집하는 것이다.

해운기업의 경우에는 전자나 후자 모두 적용이 가능한 방법이지만 연구기관의 경우에는 특정 선사의 선박이나 장비를 통해 직접 데이터를 확보하는 것은 현실적으로 무리가 있다.

선사의 경우 선박관리 조직이 선박으로부터 직접 데이터를 확보하는 사례가 있다. 선박관리회사에 의하면 선박에 별도의 센서를 부착하지 않고도 수천 건의 데이터를 수집하여 육상으로 전송하는 것이 가능하다. 요즘 대부분의 항해, 기관 장비들이 전자적으로 조정·통제되기 때문에 이미 장비 자체에 센서가 부착되어

있어 이러한 정보수집이 가능한 것이다. 운항정보는 VDR(Voyage Data Recorder)을 통해서 수집하고 기관실의 운전관련 정보는 AMS(Alarm Monitoring System)을 통해서 수집한다. 수집된 정보는 본선의 전산시스템으로 동기화하고 일정한 시간간격으로 위성송신 시스템을 통해 육상으로 전송한다.

선사가 선박에서 확보하는 정보는 대부분 특정 선박 또는 기기의 물리적인 상태나 이동에 대한 정보이기 때문에 시장분석에 사용되는 데에는 한계가 있다. 하지만 일부 선사에서는 인공지능기법을 사용한 분석을 통해 운영상의 효율을 개선하는 목적으로 활발히 사용하고 있다. 선박에서 사물인터넷으로 수집된 정보가 활용되는 영역은 주로 연료효율의 개선이나 장비유지보수체계의 개선이다.

실제 국내 대형 선사의 선박관리 조직의 실무 책임자를 인터뷰한 결과 선박에서 수집하는 정보에 대한 선사의 입장을 확인할 수 있었다.

선사에서는 선박에서 수집하는 데이터 가치를 인식하고 있다.

선박운항 관련 정보는 선박의 유지보수나 에너지 효율 개선 등에 활용될 수 있다.

주기적으로 측정되는 빅데이터 성격의 데이터는 인공지능 분석기법을 적용하여 분석됨으로써 유의미한 정보로 변환할 수 있다. 인터뷰 대상 선사의 경우 산학공동연구 형태로 인공지능 기반의 분석을 추진 중이었다.

선박으로부터 수집되는 데이터는 기업의 고유 자산이므로 외부와 공유할 수 없다.

조선소 입장에서는 이러한 선박기기운영 관련 데이터가 선박이나 장비의 개선에 매우 귀중한 자료이므로 자료를 보유한 선사에 자료를 활용한 공동연구를 제안하는 경우도 종종 있다고 한다.

특정 터미널에서 운영상의 효율성을 제고하기 위해 RFID 등 감지장치를 활용한 데이터를 수집·처리하는 사례는 있으나 선사 또는 제3자의 기관에서 간선 교통망이나 특정 물류거점에 감지기를 부착하여 빅데이터를 수집·분석하는 사례는 아직 국내에서 보고된 것이 없다. 하지만 IoT 기반의 데이터 수집과 기계학습을 활용한 분석이 보편화되면서 앞으로 사물인터넷 기반의 수집과 분석이 점점 더 활발해질 것으로 예상된다.

2. 데이터 풀(pool)의 형성

데이터 풀링(pooling)은 다수의 선사가 자사가 보유한 데이터를 공통의 플랫폼에 투입하여 분석하고 그 결과를 공유하는 것이다. 지속가능성(sustainability)을 추구하는 비영리조직인 BSR(Business for Social Responsibility) 산하의 CCWG(Clean Cargo Working Group)에서 이산화탄소 배출량, 운항거리 등의 정보를 다수의 선사로부터 수집하여 분석한 후 벤치마크를 공유하는 것이 데이터 풀링의 작은 사례이다.

풀링이 성공하려면 다음의 세 가지 조건이 충족되어야 한다. 첫째, 개별기업들이 빅데이터 분석역량에 한계를 느끼고 있어야 한다. 둘째, 관측치가 증대되면서 분석 결과물의 통계적 유의성이 높아져 가치가 증대된다는 확신이 있어야 한다. 셋째, 참여 선사들 간에 데이터 공유에 따른 경쟁관계 상 득실을 계산하지 않아야 한다.

인터뷰에 의하면 국내의 대부분의 기업들은 빅데이터 분석을 자체적으로 수행할 능력을 갖추지 못하고 있고 외부의 역량을 활용하는 측면에서도 경제적인 부담을 느끼고 있다. 따라서 첫 번째 조건의 충족은 비교적 용이한 일이 될 것이다.

두 번째 조건을 볼 때 통계적으로 관측치가 많아질수록 분석의 정확도와 신뢰도가 증가하게 된다. 특정 해운기업이 생성하는 데이터의 경우 관측치가 적어 일반화하거나 통계적인 유의성을 확보하기 어려운 경우가 많다. 오랜 기간에 걸쳐 분석에 활용될 수 있는 충분한 양의 데이터를 확보하였다고 하더라도 그 기간에 걸쳐 이루어진 기술적 진보나 여건의 변화를 감안하면 데이터의 통계적 유의성을 확보하기 어려운 경우가 많다.

예를 들어 특정 선형이 어떤 항로에서 시점에 따라 기상이나 조류의 영향을 얼마나 받는지 측정하기 위해 선박으로부터 데이터를 수집하는 경우를 가정해 보자. 그 회사가 가지고 있는 몇 척의 동형선으로부터 데이터를 확보한다고 해도 일반화가 가능한 수준에 이르지 못할 가능성이 크며 장기간에 걸쳐 데이터가 확보되었다 하더라도 데이터의 유효성에 문제가 있는 경우가 많을 것이다.

이런 경우 유용한 해결방안이 풀링(pooling)이다. 만일 많은 기업이 참여하여

데이터를 축적한다면 단기간에 많은 데이터의 축적이 가능할 것이다. 이러한 데이터 풀(pool)은 국내 기업뿐만 아니라 통합에 의한 가치의 상승을 인정하는 해외의 많은 기업들을 포함하여 이루어질 수 있다.

세 번째 조건이 가장 충족시키기 어려울 것으로 생각된다. 대형선사와 중소형 선사 간에 존재하는 데이터 양(quantity)의 불균형, 사업영역의 차이, 근린 경쟁관계 등이 제약요소로 작용할 가능성이 크기 때문이다. 이 문제를 해소하기 위해서는 정책적인 측면에서 동기부여를 위한 추가적인 고려가 필요할 것이다.

한 가지 유의해야 할 사항은 대부분의 기업들이 기업이나 선박에 관한 고유정보의 노출에 민감하다는 것이다. 따라서 민감한 정보는 수집할 때 코드화 하든지, 수집에서 제외해야 할 것이다. 또한 이러한 우려를 불식시키기 위해서 풀(pool)의 운영을 담당하는 기관을 신뢰성 있는 기관으로 선정하는 것이 중요하며 풀에 참여하는 기관과 비밀유지협약(NDA) 등을 체결하여 추가적인 신뢰를 제공할 필요가 있다.

제한적인 기업을 대상으로 한 조사이기는 하나 실무자 인터뷰를 통하여 확인한 사항은 국내의 경우 근린 경쟁관계에 대한 의식이 강하게 작용하기 때문에 공유에 대한 인센티브의 제공 등 추가적인 동기가 작용하지 않으면 자발적인 공유는 어려운 상황이다.

또한 데이터의 공유를 통한 가치의 증대를 실현하기 위해서는 실무자보다는 경영층의 동의와 의사결정이 필수적이다. 실무자들은 정책적으로 결정되지 않은 데이터의 유출에 대하여 '정보보안'의 관점에서 공유의 한계를 인식하고 있기 때문이다.

3. 개방형 데이터에 연결

공공데이터의 경우 이미 많은 데이터가 공개되어 있고 앞으로의 추세도 공개의 범위가 확대되는 방향으로 전개되고 있다. 또한 공개 시에 정부부처를 중심으로 연관 주제의 데이터를 통합해서 제공하는 것도 활발히 시도되고 있다.

정부에서는 공공데이터 포털(www.data.go.kr)을 통하여 오픈 API의 형태로

공공성을 가진 데이터를 광범위하게 제공하고 있으며, 해양수산부에서는 해운항만물류정보센터(SP-IDC)를 운영하여 해운항만통계, 선박입출항정보, 수출입물동량 등의 정보를 통합적으로 제공하고 있다. 국토교통부에서는 국가물류통합정보센터를 개설하여 해운항만통계 뿐만 아니라 항공화물통계와 내륙화물통계까지 다양한 운송모드를 망라한 통계를 종합적으로 제공하고 있다.

데이터 개방성의 증대는 우리나라뿐만 아니라 많은 나라에서 공통적으로 진행되고 있다.

송효진 등(2014)⁸¹⁾에 의하면 미국은 에너지, 교육, 재정, 공공안전, 글로벌 협력 등 5대 우선개방분야를 중심으로 공공부문의 데이터 공개가 적극적으로 이루어지도록 추진하고 있다. 법적 근거는 Electronic Freedom of Information Act이며 대부분의 지방자치단체가 SOCRATA라는 개방 플랫폼을 사용하고 있다.

영국은 2005년부터 ‘공공데이터 재활용과 규제안’을 시행하고 있으며 CIE(Common Information Environment)정책을 구체화하였다. 규제안에 의하면 공공기금이 들어간 모든 과학 연구물에 일반인의 접근을 허용하고 있으며 모든 공공 정보에 대해 정당한 이유가 없는한 재이용을 허용하고 있다.

일본은 정부나 독립행정법인, 지방자치단체가 보유하고 있는 다양하고 방대한 데이터를 기계적인 형식으로 전환하여 영리목적을 포함한 자유로운 편집 및 가공이 가능하도록 하는 등 비교적 넓은 의미에서의 공공데이터 개방을 추진하고 있다. 뿐만아니라 일본에서는 각 지방자치단체 간 오픈데이터 연계 노력도 적극적으로 이루어지고 있다.

개방데이터를 연결하는 핵심기술인 API는 Application Programming Interface의 약어로서 운영체제나 프로그래밍 언어가 제공하는 기능을 제어할 수 있게 만든 인터페이스이다. 파일, 창, 화상, 문자 등을 제어하여 서비스정보, 데이터 등을 누구나 쉽게 이용할 수 있도록 개방된 경우 이를 개방형(open) API라고 한다.

실제 운영은 개방형 데이터 플랫폼에 있는 데이터를 이용하여 개발자가 사용

81) 송효진·황성수(2014), “정부 3.0 추진에 따른 공공데이터 개방과 지방정부의 방향성 모색 : 공공데이터 법에 관한 이해와 개방사례를 중심으로”, 『한국지역정보학회지』 제17권 제2호, pp. 1~28

자 콘텐츠를 결합하여 응용프로그램을 작성하는 서비스를 제공하는 형태이다. 개방형 API라고 하더라도 데이터를 사용하기 위해서는 API 인증키를 요청해서 발급 받아야 연결하여 개발하는 것이 가능하다. 개방형 API는 주로 구글, 네이버 등 대형 웹서비스 업체에서 제공되며 이를 구현하기 위해서는 SOAP, XML, Java Script 등의 구현기술이 사용되어야 한다.

우리나라에서 개방형 API가 제공되는 예를 보면, 공공데이터 포털(www.data.co.kr)에서 도로명 주소, 국내 관광정보, 수산물 수출입정보, 부동산 상권정보 등 다양한 정보 약 2,300여 종을 제공하고 있다.

국내에 글로벌 해운시장의 분석에 활용할 수 있는 공공데이터는 매우 제한적으로 존재한다. 원천데이터로 생성되는 공공데이터가 Port-MIS와 관세청데이터 등에 불과하기 때문이다. 그나마도 국내에 국한되기 때문에 글로벌 해운시장의 분석에 활용되는 측면에서는 유용성이 크지 않다.

4. 수탁이나 공동연구를 통한 확보

UNCTAD, IMO, 선급협회 등 해운에 대한 데이터를 확보하고 있는 많은 기관이나 기구들이 도메인 지식(domain knowledge)과 분석역량의 부족으로 데이터를 처리하거나 활용하지 못하는 경우가 많이 있다. 해운연구기관이나 기업이 도메인 지식과 분석역량을 갖추고 있는 경우 해당 연구기관이나 기업은 데이터를 확보한다는 측면에서, 그리고 분석을 위탁하는 기관은 결과를 향유한다는 측면에서 상호 윈-윈 하는 입장에 놓이게 된다. UNCTAD가 한국해양수산개발원과 데이터 처리를 위한 MOU 체결을 추진하고 있는 것이 좋은 사례이다.

이러한 기관간의 협업은 자료 분석을 위탁하는 형태로도 이루어질 수 있고 두 기관 간의 공동연구로도 진행될 수 있다. 공동연구로 진행되는 경우 두 기관의 역량이 결합하여 추가적인 시너지(synergy)를 창출하는 것도 가능하다.

기업의 담당자 인터뷰를 통해 확인한 바로는 기업이 데이터의 분석을 위한 역량을 내부적으로 보유하는 것은 대기업의 경우에도 매우 어려운 일이다. 따라서 아웃소싱을 통해 분석역량을 확보하는 것이 불가피한데 역량의 아웃소싱에는

비용의 부담이 수반되는 것을 장애요인으로 인식하고 있다. 특히 해운이 장기불황으로 고통을 겪고 있는 상황에서 비용절감이 지속적으로 추진되기 때문에 필요성을 인식하고 있는 경우에도 실행에 옮겨지지 못하는 경우가 있는 것이 현실이다.

이런 경우에 연구 및 분석 역량을 보유한 기관의 활용에 정책적인 인센티브의 제공이나 연구기금의 활용 등이 결합된다면 실질적인 효과를 거둘 수 있을 것이다.

5. 외부데이터의 구매

외부의 공급기관으로부터 데이터를 구매하는 것은 가장 손쉬운 방법이지만 하나 경제적인 측면에서는 많은 기업이나 기관들이 부담을 갖는 방법이기도 하다.

국내에는 해운시장분석에 활용되는 원천데이터를 상업적 목적으로 수집하여 판매하는 기업이 거의 없으나 해외에는 여러 기업이 이러한 형태의 공급을 담당하고 있다. 해운업계에 가장 많이 알려진 공급자로는 Clarkson Research가 있으나 공급되는 데이터가 빅데이터의 성격을 가진 데이터라기보다는 전통적 데이터의 범주에 속하는 것들이다. 빅데이터 분석이라고 해도 기계학습을 활용한 분석의 단계에서는 전통적 데이터와 결합하여 입력변수로 투입되기 때문에 전통적 데이터도 그 중요성이나 활용도 측면에서는 적어도 빅데이터와 동등한 가치가 있는 것으로 볼 수 있다.

상업적으로 판매되는 데이터 중 빅데이터의 성격을 가지고 있는 것은 실시간으로 선박의 동정을 파악하는 AIS자료가 있다. 다수의 기업이 AIS데이터를 확보하는 시스템을 갖추어 놓고 데이터를 수집하여 판매하고 있는데 대표적인 기업으로 IHS Markit과 FleetMon을 들 수 있다.

AIS데이터는 물리적인 선박의 이동 현황에 관한 자료이기 때문에 이를 상업적 의사결정에 활용하기 위해서는 도메인 지식에 기초한 창의적인 접근이 필수적이다.

전통적인 데이터의 구매에는 큰 비용이 소요되지 않기 때문에 비교적 구득이 용이하나 국내선사 중에 Clarkson데이터나 Baltic Exchange의 시계열 자료를

유료로 구매하고 있는 선사는 많지 않은 것으로 파악되었다.

예를 들어 Baltic Exchange는 회원가입을 통해 운임시장에 대한 자료 즉 운임지수와 운임선물시장의 자료를 확보할 수 있는데 국내선사 중 Baltic Exchange의 회원자격을 확보한 선사는 현대상선을 비롯한 극소수인 것으로 밝혀졌다. FFA시장에 대한 시계열 자료의 경우 FFA가 가진 가격발견기능 때문에 중단기 시황예측에 매우 중요한 자료임에도 확보한 선사가 제한적인 것을 보면 국내선사의 정량적 분석의 수준이 그리 높지 않음을 유추할 수 있다.

Clarkson의 경우 온라인 데이터의 사용이 연간 USD6,650+VAT이며 Baltic의 경우 선도가격의 시계열 자료를 포함한 경우 연간 GBP4,420+VAT로 파악되었다.

AIS데이터의 경우 업체에 따라 금액의 편차가 크고 데이터가 고객 맞춤형으로 제공되기 때문에 파악이 쉽지 않으나 전통적 데이터보다는 고가일 것으로 추정된다.

6. 웹탐색

인터넷으로 연결된 외부 사이트에서 데이터를 수집하는 것으로 수집을 위해서는 수집용 도구에 해당하는 크롤링엔진(crawling engine)을 사용한다. 수집하는 형태는 다양하나 웹사이트에 게재되어 있는 웹 로그(블로그)나 이벤트 로그를 수집하는 것이 주종을 이룬다. 해운서비스는 일반 상품과 다르기 때문에 이벤트 로그 보다는 웹 콘텐츠를 수집하여 분석하는 것이 유용할 것이다. 수집된 자료는 텍스트 마이닝이나 감성분석의 과정을 거쳐 거래처의 평판 및 신용분석, 시황국면의 판단 등에 활용될 수 있다.

분석을 위해 탐색하는 웹 사이트의 범위가 큰 것이 일반적이므로 기존의 관계형 데이터베이스 처리기술로는 한계가 있어 Hadoop 등 빅데이터 처리 시스템의 활용이 필요하다.

국내 해운기업 중 웹탐색을 통해서 비정형의 텍스트 데이터를 분석하여 시황의 판단이나 거래처의 신용판단에 활용하는 사례는 아직 보고되지 않고 있다.

하지만 텍스트 데이터와 감성분석이 타 분야의 마케팅 분야에서는 활발히 사용되고 있으므로 해운업에 있어서도 활용의 가능성은 크다고 볼 수 있다.

이를 실현하기 위해서는 첫째, 해결하고자 하는 문제가 명확히 정의되어야 하며 둘째, 데이터가 확보되어야 하고 셋째, 분석역량이 갖추어져 있어야 한다.

데이터의 확보 측면에서는 대부분의 해운 관련 보고서나 뉴스가 유료화 되어 있기 때문에 데이터의 구매가 필요할 것으로 판단된다.

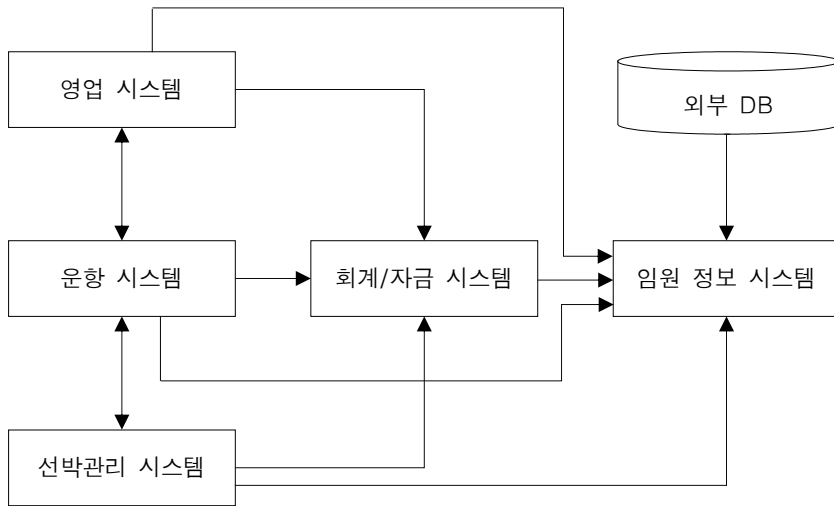
분석역량은 내부적으로 확보하는 것이 어려우므로 전문분석기관이나 학계의 역량을 확보할 필요가 있다. 개별기업의 이러한 경제적인 부담을 경감하는 차원에서 이러한 분석역량을 공공역량화해서 분석결과를 공공서비스로 국내기업에 제공하는 것이 좋은 대안이 될 수 있다.

7. 이해당사자 간의 교환

동일한 영역에서 사업을 영위하는 기업 또는 다른 영역에 있더라도 통합을 통한 가치의 상승을 기대할 수 있는 이종데이터를 보유한 기업/기관 간에는 교환이 바람직한 대안이 될 수 있다. 교환의 대상이 되는 데이터는 기업이 통합전산 시스템에서 생성되는 자료 또는 사물인터넷 기반으로 확보하는 자료이다.

IT 시스템 역량을 주요 경쟁역량으로 삼고 있는 컨테이너 선사들은 물론이고 벌크 선사들도 규모가 일정 수준 이상이면 회계시스템에 추가하여 영업시스템, 운항시스템, 선박관리시스템 등의 통합시스템을 구축하여 운영하고 있다.

〈그림 4-2〉 선박회사의 전산시스템 운영개요도



자료 : 저자정리

이러한 통합시스템으로부터 산출되는 데이터는 소위 분석 시스템으로 전달되고, 임원정보시스템(EIS: Executive Information System)으로 가공되어 의사결정을 지원하게 된다.

시장분석의 관점에서 우리가 주목해야 할 부분은 개별기업의 차원에서 축적되는 데이터가 아니라 공유를 통해 가치가 급격히 증가하는 데이터이다. 선박의 영업 및 운항과 관련하여 개별기업이 축적하는 자료는 그 자체로는 관측치가 너무 적어 통계적인 분석을 하더라도 유의성을 갖기가 어렵지만 많은 기업의 자료가 통합되면 통계적인 추세나 특성을 파악하는 것이 유의미해진다. 예를 들어 극동에서 동남아로 항해하는 선박의 항해 일정, 선속, 연료소모량 등이 연중 측정되고 이러한 데이터가 대규모로 통합된다면 해류 또는 기상조건의 변화에 따른 연료효율의 계절적 변동성을 정확하게 파악하여 항차예산의 정확도를 높이는 데 기여할 수 있을 것이다.

해운기업에서 사물인터넷 기반으로 수집하는 자료는 제 1절에서 다루어진 바와 같이 주로 선박에서 수집하는 자료이다. 최근 기업들은 이러한 자료의 가치를 인식하고 기존의 센서 뿐만이 아니라 추가적인 센서를 비용을 들여 부착해서 수

집하는 추세이다. 또한 수집된 자료를 인공지능 등 분석도구를 활용해서 선박운영의 효율성을 제고하려는 움직임을 보이고 있다. 기업은 이러한 분석 역량을 산학 협력 또는 독립 연구기관을 통해 확보하는 것으로 알려져 있다.

8. 추진방안

위에서 일곱 가지의 다양한 확보방안에 대하여 살펴보았다.

사물인터넷 기반으로 확보된 자료는 개별기업의 차원에서 활용하는 것보다는 데이터 풀이나 교환 등의 방법으로 관측치를 확대하여 활용하는 것이 필요하다.

데이터 풀의 경우 정부차원의 정책적 지원을 통하여 공유를 촉진하고 분석역량은 공공역량화 함으로써 현실적인 대안으로 승화시킬 수 있을 것이다.

개방형 데이터는 우리나라뿐만 아니라 해외까지 확대함으로써 그 유용성을 제고하는 것이 필수적이다.

수탁이나 공동연구는 한국해양수산개발원 등이 공공역량화의 주체로서 기여할 수 있는 확보방법에 속한다. 이를 위해서는 한국해양수산개발원 등이 내부적으로 역량을 확보함과 동시에 다양한 형태의 학연연구를 통해 역량을 확보할 필요가 있다.

외부데이터의 구매도 공공역량화를 통해 중복구매를 막고 기업의 부담을 경감해 줄 수 있는 방법이므로 다른 확보방법과 연계하는 방안이 강구되어야 한다.

웹탐색 또한 해운관련 데이터의 유료화 경향을 고려하면 외국데이터의 구매와 마찬가지로 공공화의 관점에서 접근할 필요가 있고 분석 또한 공공역량으로 수행하는 것이 바람직할 것이다.

한국해양수산개발원에서 수행할 경우 해운시장분석을 담당하는 조직과 통계분석을 담당하는 조직이 협업하고 외부의 역량을 효과적으로 결합할 필요가 있다.

공적연구기관이 상업적 의사결정을 다루는 것에 대하여 적절성 측면의 비판이 있을 수 있으나 현재 해운 및 유관 산업의 어려운 현실과 향후 업계의 역량 개발 속도를 고려할 때 앞으로 상당기간은 정부 출연연구기관이 주도권을 잡고

데이터 확보와 분석을 수행하는 것이 바람직할 것으로 판단된다.

끝으로 각 확보방법별로 한국해양수산개발원의 입장에서 데이터를 확보할 수 있는 가능성, 해운업의 중요의사결정에 대한 지원 측면에서의 효과성, 확보의 경제성 측면에서 요약한다.

〈표 4-3〉 확보 방법별 특성

	확보가능성	효과성	경제성
사물인터넷기반의 수집	낮음	낮음	낮음
데이터 풀의 형성	중간	낮음	높음
개방형데이터에 연결	중간	중간	높음
수탁이나 공동연구를 통한 확보	중간	낮음	높음
외부데이터의 구매	높음	높음	낮음
웹탐색	중간	중간	중간
이해당사자 간의 교환	낮음	낮음	높음

제5장

결론 및 정책제안 <<

제1절 결론

1. 연구내용의 요약

해운업의 의사결정 중 중요한 것으로 간주되는 영역이 사업구조의 결정에 대한 의사결정과 선복⁸²⁾의 확보 및 처분과 관련된 의사결정으로 파악되었다. 사업구조에 대한 의사결정이든 선복의 매매(trading)에 대한 의사결정이든 중요한 것은 정보에 기초한 합리적인 의사결정이어야 한다는 것이다.

해운기업의 의사결정 관행을 보면 과학적인 요소가 상당히 결여되어 있는 것을 알 수 있다. 많은 기업이 데이터를 과학적으로 분석하기 보다는 외부의 시장 보고서를 적절히 조합한 것을 의사결정의 준거로 삼고 있다. 약 58%의 응답자가 시계열 분석을 중요한 분석기법으로 꼽고 있기는 하나 실제 정량적 분석에 대한 신뢰도는 그다지 높지 않아 유용하다고 판단한 응답자는 전체의 27%에 불과했다.

분석적인 접근이 제한적으로 활용된 데에는 다양한 이유가 있겠으나 과거의 성과가 좋지 못해 신뢰성을 확보하지 못한 것이 주요 원인 중의 하나이다.⁸³⁾

4차 산업혁명시대를 맞아 데이터 분석을 둘러싼 환경은 급격한 변화의 시기를 맞고 있다. IoT기술을 기반으로 한 빅데이터의 수집이 데이터의 새로운 지평을 열고 있으며 기계학습이나 인공지능을 활용한 분석도구들이 기존의 정량적인 분석기법들을 능가하는 성과를 보여 새로운 접근으로 인식되고 있다.

⁸²⁾ 이 보고서에서 선복은 선박의 구매와 용선을 통해 확보된 수송능력을 총체적으로 일컫는 용어로 사용됨. 따라서 '선박'의 매매와는 다른 의미를 갖는 개념이다.

⁸³⁾ Stopford, M. (2009) Maritime Economics, Routledge, Abingdon pp.697-698

이 연구는 이러한 환경의 변화를 해운시장 분석이라는 관점에서 어떻게 수용할 것인지를 다루는 연구 중 첫 번째 단계의 연구이다. 연구의 주요 내용은 해운시장분석의 현황을 파악하는 것과 2단계에서 분석기법을 적용하기 전에 데이터의 소재와 구조를 파악하고, 데이터 확보 방안을 강구하는 것이다.

현황의 파악은 설문조사를 통해서 이루어졌으며 설문조사는 해운의사결정의 유형, 의사결정의 준거, 해운시장분석, 활용되는 데이터, 빅데이터 분석에 대한 인지도를 내용으로 하고 있다.

해운의사결정은 사업구조의 결정과 선복거래의 시점/기간에 대한 의사결정의 비중이 높은 것으로 조사되었다. 운영상의 효율성을 개선하기 위한 의사결정 보다는 전략적이고 시장위험(market risk)관리에 직접적인 영향을 주는 영역이 중요한 의사결정으로 간주되는 것을 보여주고 있다.

의사결정의 준거에 대해서는 외부에 전적으로 의존하기 보다는 내부적인 통합과 조정의 과정을 거친 내부보고서에 대한 의존도가 큰 것으로 나타났다.

해운시장 분석에는 시계열 분석이 주로 활용되고 있으며 분석을 위한 조직은 3-5명의 소규모 조직이 높은 비중을 차지했다. 내부분석 결과는 대체로 잘 수용되는 편이나 최고 의사결정권자의 성향에 따라 크게 변화한다는 응답이 많아 권한 있는 소수에 의해 의사결정이 좌우되는 현실을 드러내고 있다.

데이터의 원천으로는 외부의 시황분석 보고서나 뉴스에 대한 의존도가 큰 것으로 확인되었다.

빅데이터에 대한 인지도는 그다지 높지 않으나 빅데이터 분석이 해운의사결정에 긍정적인 영향을 미칠 것으로 보는 응답자가 전체의 70%를 넘었다. 하지만 실제 빅데이터를 수집, 분석하고 의사결정에 활용하는 경우는 아주 제한적인 것으로 파악되었다.

빅데이터의 파악에 대해서는 1차 자료(primary data)의 원천을 파악하는 것을 대상으로 하였다. 국내에서 해운시장분석과 관련된 자료가 1차 자료의 형태로 생성되는 경우는 많지 않다. 공공데이터의 경우 Port-MIS에서 물동량의 물리적인 흐름을 포착하는 것이나 관세청이 무역의 흐름을 수출입 신고액으로 파악하는 것이 이에 해당하는 것이며 여타의 기관에서 다루어지는 대부분의 자료들은 이를 정리하거나 가공한 것에 해당한다.

일반기업에서는 선박의 운항정보를 사물인터넷으로 선박에서 직접 수집하는 것으로 파악되었다.

실제 시장분석에 효과적으로 활용할 수 있는 대부분의 데이터는 데이터를 유상으로 공급하는 상업적 기관으로부터 확보할 수 있다. 이들 대부분은 전통적 데이터의 범주에 해당하는 것이고 빅데이터로 분류될 수 있는 것은 선박의 자동 식별시스템(Automatic Identification System)으로 확보되는 데이터이다. 이 연구에서는 Clarkson Research, Baltic Exchange, IHS Markit 등에서 제공하는 전통적 데이터의 구조를 탐구했으며 AIS 데이터는 IHS Markit, FleetMon 등을 연구하였다.

빅데이터의 확보는 아래와 같은 다양한 방법으로 이루어진다.

사물인터넷 기반의 수집

데이터 풀(pool)

개방형 데이터에 연결

수탁·공동연구 목적의 공유

웹 탐색

교환

구매

위의 대안 중 시장분석을 공적인 역량으로서 수행하는 한국해양수산개발원의 입장에서 가장 현실적인 대안은 데이터 공급업체로부터 구매하는 것이다. 대부분의 시계열 자료나 AIS자료는 상업적인 공급기관으로부터 구매할 수 있다.

한편 UNCTAD, IMO, 선급 등 데이터는 확보하고 있으나 분석역량이 충분하지 않은 조직에 분석역량을 제공함으로써 데이터를 무상으로 확보하는 것은 가장 경제적이고 창의적인 접근이 될 것이다.

또한 대부분의 데이터가 관측치가 증가할수록 입력변수의 가치가 증대되므로 많은 기업/기관들의 데이터를 풀(pool)의 형태로 통합함으로써 향상된 분석결과를 산출할 수 있다. 연구기관은 이 통합의 주체가 됨으로써 데이터를 무상으로 확보할 수 있을 것이다.

이외에 개방형 데이터에 연결하거나 웹 탐색, 사물인터넷 기술의 적용을 통한 1차 자료의 수집은 필요에 따라 실행될 수 있을 것이다. 웹 탐색이나 사물인터넷

넷 기반의 자료들은 엄밀한 의미의 빅데이터가 될 정도로 방대한 크기가 될 수 있어 분산저장, 분산처리 기술의 적용이 필요하게 된다.

한국해양수산개발원의 경우 현재로서는 1차 자료를 수집하지 않아 자료의 교환은 현실적인 대안이 되지 못한다. 하지만 웹 탐색이나 사물인터넷을 통한 직접 수집의 사례가 생기는 경우 이 데이터를 다른 기업이나 기관이 보유한 데이터와 교환함으로써 상호간에 데이터 보유가치의 증대를 기대할 수 있을 것이다.

2. 연구의 기대효과

1) 이 연구의 효과

이 연구는 3단계로 진행되는 ‘빅데이터를 활용한 시장분석 연구’의 제 1단계 연구이다. 연구는 기업의사결정의 초점파악, 기존 시장분석의 현황, 빅데이터에 대한 인지도파악, 데이터의 소재 및 구조파악 그리고 데이터 확보방안의 마련을 내용으로 하고 있다.

이 연구는 2단계 모델 연구의 토대라는 측면에서 가치가 있는 연구이다. 모델을 연구하기 위해서는 의사결정 문제가 명확하게 정의되어야 하며 입력변수에 해당하는 데이터의 확보가 선행되어야 한다.

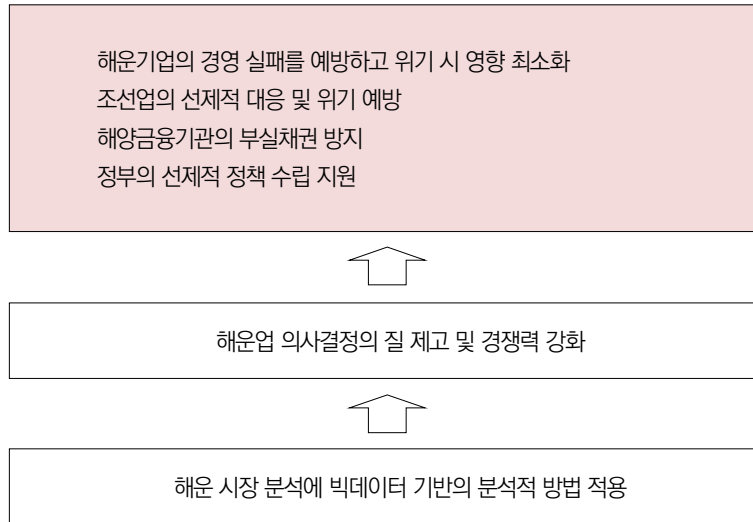
기존의 빅데이터에 대한 연구는 데이터 플랫폼에 대한 연구가 주류를 이루었기 때문에 문제해결을 위한 모델의 적용에 초점을 맞추지 못한 한계가 있었다. 이 연구에서는 해운의 의사결정과 직결되는 의사결정 지원모델에 대한 입력변수로서 빅데이터와 전통적 데이터가 다루어졌으며 중요한 데이터 원천에 대해서는 데이터의 구조가 명확히 파악되어 향후 모델 연구에서 실질적으로 활용될 수 있도록 하였다.

2) 3단계 연구의 기여

3단계 연구가 완성되고 의사결정 모델의 산출물이 제시될 경우 해운의사결정의 질이 높아지는 효과를 기대할 수 있다. 또한 2장 제 1절에서 제시된 산업의 연관 구조를 고려할 때 해운업 의사결정 역량의 강화는 조선, 선박금융, 정부 등 관련 의사결정 주체의 역량강화로 이어져 산업의 건전성을 제고하고 궁극적

으로는 산업의 실패를 예방하는 효과를 가져올 것으로 기대된다.

〈그림 5-1〉 3단계 연구의 기여



제2절 정책제안

1. 100대 국정과제와 빅데이터를 활용한 해운시장분석 연구

1) 해운·조선 상생을 통한 해운강국 건설

현 정부는 골고루 잘 사는 균형발전전략의 하나로 해운·조선 상생을 통한 해운강국 건설이라는 과제를 제시하였는데 이 과제는 이번 연구가 가장 직접적으로 기여할 수 있는 영역이다.

우리나라의 해운업은 기업, 정부, 금융기관 등 관련 주체들에 의한 누적된 의사결정으로 말미암아 불황기마다 산업전체가 심각한 타격을 받아왔다. 데이터 분석에 기반한 해운시장의 분석은 의사결정의 질을 높여 반복되는 해운업 실패를 예방하는데 기여할 것으로 기대된다.

또한 조선업은 해운을 수요의 원천으로 하는 사업으로서 해운의 건전한 발전이 없이는 성장을 도모할 수 없다. 뿐만 아니라 해운시장분석 자체가 조선시장의 수요분석의 핵심이 되므로 이번 연구의 성과가 조선업의 발전과 직접적인 관련이 있는 것으로 판단된다.

2) 소프트웨어 강국, ICT 르네상스로 4차 산업혁명 선도 기반 구축

해운시장분석 연구에는 전통적 데이터뿐만 아니라 사물인터넷(IoT) 기반으로 수집된 빅데이터가 광범위하게 활용된다. 또한 데이터의 분석단계에서는 기계학습(machine learning)이나 인공지능(artificial intelligence)을 활용한 분석도구들이 다양하게 활용된다.

이러한 직접적인 요구와 광범위한 적용이 4차 산업혁명을 이끄는 핵심적인 기술적 역량의 발전을 촉진할 것으로 예상된다.

3) 주력산업 경쟁력 제고로 산업경제의 활력회복

정부에서 제시한 이 정책과제는 과학기술이 선도하는 4차 산업혁명이라는 전략에 속한 과제이다. 해운과 조선이 과학기술 발전과 직접적인 관련성이 있는 것은 아니나 해운은 한진해운 파산의 교훈에서 배웠듯이 수출 주도형 경제에서 우리 제품의 수출 경쟁력을 결정한다는 의미에서 핵심 주력 산업의 위치를 점하는 것이 당연하며 조선 또한 다수의 협력업체와 기자재 산업을 거느린 핵심제조업으로서 국가 주력 산업으로서 자리매김하고 있다.

본 연구는 해운과 조선의 사업경쟁력확보를 직접적으로 지원함으로써 산업경제의 활력을 회복하고 고용을 유지 또는 확대하는데 기여할 것이다.

4) 중소기업의 튼튼한 성장환경 구축

현재 선주협회의 회원사 169개 중 대기업으로 분류되는 회사는 현대상선, 현대글로벌비스 등 소수에 불과하며 나머지는 모두 중소기업의 범주에 들어가는 해운사이다. 이들 중소선사는 기업의 안정성이 크지 않아 2009년 이후 선주협회 회원사에서 약 100개 회사가 소멸되고 같은 수만큼의 회사가 신설되었다.⁸⁴⁾

중소해운기업의 경우 시장분석을 위한 조직적 역량을 갖추는 것이 현실적으

로 용이하지 않다. 또한 데이터의 확보에 있어서도 상당한 제약을 갖게 된다. 빅데이터를 활용한 해운시장 분석이 공공 역량으로 개발되고 의사결정을 지원하는 서비스가 제공된다면 중소 해운기업의 경쟁력 강화에 실질적으로 기여할 수 있을 것이다.

2. 정책제안

1) 시장분석의 공공역량화

해운은 수급구조가 매우 복잡하고, 경기의 변동성이 다른 산업보다 크며, 사업구조나 노출의 조정에 비교적 장기간이 소요되기 때문에 시장분석을 통한 예측과 판단 역량을 확보하는 것이 매우 중요하다.

그럼에도 불구하고 우리나라에서는 해운업체가 시장분석역량을 제대로 갖추지 못한 것으로 파악되고 있다. 통상적으로 시장분석역량은 기업 내부와 3자 제공 역량의 결합으로 받아들여진다. 기업 내부의 역량을 보면 대기업의 경우 인적자원을 할당할 수 있는 조직적 역량은 있으나 기업 경영층의 판단에 조직의 운명이 좌우되는 경향이 있어 국내에 내부역량의 연속성이 유지되는 해운기업은 거의 없다. 중소기업의 경우 대체로 영세한 규모이고 인적자원의 풀(pool)도 크지 않기 때문에 조직적인 역량을 갖추고 있지 못하는 것이 현실이다.

제 3자 제공역량의 경우 영국과 같은 해운 선진국에서는 브로커 기반의 연구기관과 독립적인 연구기관이 고르게 발달되어 있는데 이는 그들이 창출한 서비스를 판매할 시장의 규모가 크고 중개업도 조직적으로 이를 감당할 수 있을 정도로 발전해 있기 때문이다.

국내의 경우 해운중개업은 대부분 영세기업의 형태로 유지되고 있어 해운시장분석을 위한 전담조직을 구성할 만큼의 조직적 역량을 보유하고 있지 않다. 또한 국내시장의 경우 시장규모가 작아 영국의 MSI와 같은 독립적인 연구기관도 발달하지 못했다. 국내 연구기관의 글로벌 역량이 커서 해외로 진출할 수 있었으면 사정이 달랐을 것이다. 아직 국내 연구기관으로서 세계시장을 대상으로

84) 선주협회, 한경해양산업 CEO 포럼 발표자료, 2017.10.24

연구 성과를 판매할 수 있는 기업은 비단 해운뿐만 아니라 다른 영역에도 극히 제한적으로 존재한다.

이러한 상황을 고려할 때 적어도 우리 해운산업이 역량을 축적해서 일정 수준의 성과를 낼 때까지 시장분석 역량을 공공역량으로 개발해서 필요한 서비스를 업계에 제공하는 것은 해운업의 건전한 발전과 위기의 예방이라는 차원에서 매우 의미 있는 일이다.

공적인 연구업무를 수행하는 기관이 매우 제한적이므로 공공역량의 개발주체를 선정하는 문제에는 선택의 여지가 별로 없다. 한국해양수산개발원의 경우 정책연구기관이고 해운시장분석이 민간영역이라는 괴리가 존재하나 정책연구기관의 외연을 확장하여 대응하는 것이 바람직할 것으로 판단된다.

2) 인적역량에 대한 투자

해운시장분석이 효과적으로 이루어지기 위해서는 다양한 인적역량의 결합이 필요하다. 우선 해운업이라는 업에 대한 지식(domain knowledge)이 전제되어야 한다. 아무리 분석이나 IT능력이 탁월하다 하더라도 해운업을 정확히 이해하지 못하면 문제를 정의하고 변수를 정확히 설정할 수 없다.

두 번째는 분석역량이다. 해운시장의 분석에는 다양한 기법이 활용된다. 일반적인 계량경제학적 분석의 역량은 비교적 확보하기 용이하겠지만 이러한 분석은 해운시장에서 그 유용성이 도전받고 있는 실정이고 새로운 형태의 분석인 빅데이터 분석(big data analytics)이나 기계학습 모형을 활용한 분석은 인력의 수급측면에서 상당한 애로를 겪고 있는 것으로 알려져 있다.

결과적으로 도메인 지식과 분석능력의 융합능력을 가진 인력이 체계적으로 양성되어야 한다. 이를 위해서는 기업차원에서 인력양성에 선제적인 투자가 이루어져야 하며 정책당국은 교육과정의 마련이나 교육에 대한 인센티브제공 등의 장려 정책으로 융·복합인력이 단기에 효과적으로 확보되도록 해야 할 것이다.

3) 데이터 확보 및 개방

데이터는 흔히 원유에 비유된다. 가공을 거쳐야 유용성이 확보된다는 점과 궁극적인 가치의 원천이 된다는 측면에서 적절한 비유라 할 수 있다.

국내에서 해운시장의 분석에 활용되는 원천데이터(primary data)의 확보는 매우 제한적으로 이루어지기 때문에 대부분의 데이터는 외국에서 구매를 통해 확보하게 되는데 이 경우 많은 투자가 필요하다.

우리나라도 원천데이터의 확보에 자원을 투입해서 원천데이터를 추가로 확보할 필요가 있다. 예를 들어 국내에는 AIS데이터를 직접 수집해서 판매 또는 제공하는 업체가 없다. AIS데이터는 해사안전뿐만 아니라 예측이나 판단모형의 입력변수로서 큰 가치를 지니고 있으므로 정부차원에서 육성 또는 지원함으로써 내부역량화 할 필요가 있다.

확보차원에서 또 하나의 중요한 이슈는 공공데이터의 개방이다. 4차 산업혁명의 진행과 더불어 데이터는 점점 더 공공재의 성격이 강해지고 있어 많은 국가에서 세관데이터 등 정부가 보유하고 있는 데이터를 공공데이터로 개방하고 있다. 미국의 경우에도 관세청 자료를 일정 자격요건을 갖춘 민간 기업에 개방한다. 우리나라의 경우 시장분석에 활용될 수 있는 원천 데이터의 종류와 양에 제한이 있다고 하더라도 관련 제도와 법령을 정비해서 개방의 속도를 높여나가는 것이 매우 중요하다.

참고문헌 《

〈국내 문헌〉

- 국가기술표준원 (2015), 『2015 표준기반 R&D 로드맵 스마트SCM』, p.9.
- 김성희·정병호·김재경 (2001), 『의사결정분석 및 응용(개정판)』, 영지문화사
- 김용대·조광현 (2013), “빅데이터와 통계학”, 『한국데이터정보과학회지』, pp. 959-974.
- 문화체육관광부 (2012), “정부 SNS 운영 효과 분석 및 활성화 방안”, 최종보고 발표자료.
- 미래창조과학부 (2014), 『빅데이터 활용 단계별 업무절차 및 기술 활용 매뉴얼(ver 1.0)』, pp. 33-34.
- 배혜림 (2015), “프로세스 마이닝을 이용한 생산 운영 빅데이터 분석 동향 및 사례”, 『e 매거진』, 제 22권 제 2호, pp. 36-42.
- 서강민·김종훈 (2016), “해운, 2차 치킨게임의 서막 - 현대상선, 용선료 인하 이후 생존가능성은?”, 『한국기업평가』.
- 선주협회, 『한경해양산업 CEO 포럼 발표자료』, 2017.10.24
- 송효진·황성수(2014), “정부 3.0 추진에 따른 공공데이터 개방과 지방정부의 방향성 모색 : 공공데이터 법에 관한 이해와 개방사례를 중심으로”, 『한국지역정보학회지』 제17권 제2호, pp. 1~28
- Stopford, M. (2015), 『해운경제학』 (양창호 외 3인 역).
- 신승식·김수엽 (2000), 『해운·항만 물류정보화를 위한 기반조성연구』, 한국해양수산개발원.
- IBM 기업가치연구소 (2012), 『분석: 빅데이터의 현실적인 활용』, p. 4.
- IDG Korea (2016), 『머신러닝 입문 가이드』, pp. 8-9.
- LG경제연구원 (2015), 『진화하는 인공지능 또 한번의 산업 혁명』, p. 4.

- 임종관·김우호·고병욱 (2010), “벡터자기상관 모형을 이용한 건화물선 시장 분석”, 『해운물류연구』, 제64호.
- 정보통신기술진흥센터 (2016), 『인공지능 기술 동향 및 발전 방향』, p. 14.
- 정봉민 (2015), 『해운경제학』, 블루&노트
- (주)세아이텍, “시각화를 활용한 빅데이터 분석과 공공데이터 개방사례”, 발표자료 활용, 2014.
- 주해종·김혜선·김형로 (2017), 『빅데이터 기획 및 분석』, 크라운출판사, pp. 34-36.
- 조용준·박용진·임종선 (2016), 『수산업의 빅데이터 기반구축방향』, 수산경제연구원.
- 최종희·박상우·이호춘·최지연·이윤정·송민호 (2013), 『해양수산 정보의 체계적인 관리 및 제공방안 연구』, 한국해양수산개발원.
- KMI (2017), 『The Ocean』, Vol.08.
- 편집부 (2016), 『딥러닝과 빅데이터 산업동향』, 하연.
- 한국정보화진흥원 (2017), 『2016년 BIGDATA 시장현황 조사』, p. 14.
- 해양수산부, 『해양수산 빅데이터 추진계획(안)』, 2016. 10

〈국외 문헌〉

- Alizadeh, A. H., R. O. Adland and S. Koekebakker (2007), “Predictive Power and Unbiasedness of Implied Forward Charter Rates”, *Journal of Forecasting*, 26, pp. 385-403.
- Batchelor, R., A. Alizadeh and I. Visvikis (2007), “Forecasting Spot and Forward Prices in the International Freight Market”, *International Journal of Forecasting*, 23, pp. 101-114.
- Beenstock, M. and A. Vergottis (1993), *Econometric Modelling of World Shipping*, Chapman & Hall.
- Cerchiello, P. and P. Giudici (2015), “Big Data Analysis for Financial Risk Management”, *Journal of Big Data*.

- Dikos G., H. Marcus, M. P. Papadatos and V. Papakonstantinou (2006), "Nivers Lines: A System-Dynamics Approach to Tanker Freight Modeling", *Interface*, Vol. 36, No. 4, pp. 326-341.
- Ducruet, C. and F. Zaidi (2012), "Maritime Constellations: a Complex Network Approach to Shipping and Ports", *Maritime Policy & Management*, Vol. 39, No. 2.
- Economist Intelligence Unit, "Retail banks and big data: Big data as the key to better risk management", *The Economist*, 2014
- Feng Z, Zhu Y (2016), "A Survey on Trajectory data Mining Techniques and Applications", *IEEE Access*
- G.Miner, D.Delen, J.Elder, A.Fast, T.Hill, R.Nisbet, Elsevier (2012), *Practical text mining and statistical analysis for non-structured text data applications*, chap. 2. p.31
- Goerlandt, F. and P. Kujala (2011), "Traffic Simulation Based Ship Collision Probability Modeling", *Reliability Engineering & System Safety*, Vol. 96, No. 1, pp. 91-107.
- Jeffrey L. Solka (2008), *Text Data Mining: Theory and Methods*
- Kavussanos, M. G. and N. K. Nomikos (2003), "Price Discovery, Causality and Forecasting in the Freight Futures Market", *Review of Derivatives Research*, 6, pp. 203-230.
- Kavussanos, M. G., I. D. Visvikis and D. Menachof (2004), "The Unbiasedness Hypothesis in the Freight Forward Market: Evidence from Cointegration Tests", *Review of Derivatives Research*, 7, pp. 241-266.
- Keen, P. G. W. and M. S. Scott Morton (1978), "Decision Support Systems: An Organizational Perspective", Reading, MA: Addison-Wesley, Inc.
- Kim, S. and D. H. Shin (2016), "Forecasting Short-term Air Passenger Demand using Big Data from Search Engine Queries", *Automation*

- in Construction*, Vol. 70, pp. 98-108.
- Kisa, R. "Knowledge management", 2014
- Li, J. and M. G. Parsons (1997), "Forecasting Tanker Freight Rate Using Neural Networks", *Maritime Policy and Management*, Vol. 24, No. 1, pp. 9-30.
- Li, J., Z. Xu, H. Xu, L. Tang and L. Yu (2017), "Forecasting Oil Price Trends with Sentiment of Online News Articles", *Asia-Pacific Journal of Operational Research*, Vol. 34, No. 02.
- Lun, V. H. and M. Browne (2009), "Fleet Mix in Container Shipping Operations", *International Journal of Shipping and Transport Logistics*, Vol. 1, No. 2.
- Lyridis, D., P. Zacharioudakis, P. Mitrou and A. Mylonas (2004), "Forecasting Tanker Market Using Artificial Neural Networks", *Maritime Economics and Logistics*, pp. 93-108.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, A. H. (2011), *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute, New York.
- Nguyen, T., L. Zhou, V. Spiegler, P. Leromonachou and Y. Lin (2017), "Big Data Analytics in Supply Chain Management: A state-of-the-art literature review", *Computers and Operations Research*, *In press*.
- Perera, L. P. and B. Mo (2016), "Machine Intelligence for Energy Efficient Ships : A Big Data Solution", *Maritime Technology and Engineering, CRC press*, p. 143.
- Randers, J. and U. Goluke (2007), "Forecasting Turning Points in Shipping Freight Rates: Lessons from 30 Years of Practical Effort", *System Dynamics Review*, Vol. 23, No. 2/3, pp. 253-284.
- Rodseth, O. J., L. P. Perera and B. Mo (2016), "Big Data in Shipping -

- Challenges and Opportunities”, In proceedings of *the 15th International Conference on Computer Applications and Information Technology in the Maritime Industries* (COMPIT 2016), pp. 361-373.
- Sage, A. P., *Decision Support Systems Engineering*, 1991
- Shen, S., H. Jiang and T. Zhang (2012), “Stock Market Forecasting using Machine Learning Algorithm”, *Semantic Scholar*.
- Singh, A., D. Jain, I. Mehta, J. Mitra and S. Agrawal (2017), “Application of Big Data in Supply Chain Management”, *Materials Today: Proceeding*, Vol. 4, No. 2, Part A, pp. 1106-1115.
- Stateczny, A. and M. Wlodarczyk-Sielicka (2014), “Self-organizing Artificial Neural Networks into Hydrographic Big Data Reduction Process”, *Rough Sets and Intelligent Systems Paradigms*, Chapter 34, Springer.
- Stopford, M. (2009) *Maritime Economics*, Routledge, Abingdon
- Tsao, Y. (2017), “Managing Default Risk under Trade Credit: Who should implement Big-Data analytics in supply chains?”, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 106, pp. 276-293.
- Wang, G., A. Gunasekaran, E. Ngai and T. Papadopoulos (2016), “Big Data Analytics in Logistics and Supply Chain Management: Certain investigations for research and applications”, *International Journal of Production Economics*, Vol. 176, pp. 98-110.
- Wang, H., O. L. Osen, G. Li, W. Li, H. Dai and W. Zeng (2015), “Big Data and Industrial Internet of Things for the Maritime Industry in Northwestern Norway”, *TENCON 2015 - 2015 IEEE Region 10 Conference*.
- Wimmer, H. and R. Rada (2018), “Applying Artificial Intelligence to Financial Investing”, *Encyclopedia of Information Science and*

Technology, ed. 4, Ch. 1.

Willems, N., H. V. D. Wetering and J. V. Wijk (2009), "Visualization of Vessel Movements", *Computer Graphics Forum*, Vol. 28, No. 3, pp. 959-966.

Witkowski, K. (2017), "Internet of Things, Big Data, Industry 4.0 - Innovative Solutions in Logistics and Supply Chains Management", *Procedia Engineering*, Vol. 182, pp. 763-769.

Yun, H., S. Lim and K. Lee (2017), "The Value of Options for Time Charterparty Extension : an Artificial Neural Networks(ANN) Approach," *Maritime Policy & Management*.

〈인터넷 자료〉

장동인 (2015), CIO Summit 2015, http://www.cuvix.co.kr/cio_summit2015/Images/Track1-1_CIO_Summit_2015.pdf, 검색일 : 2017. 08. 21

Thor Olavsrud (2017), 전문가들이 말하는 2017년 빅데이터·분석 전망 15선, <http://www.ciokorea.com/news/33014#csidxc14ca138a97d600b555c41c8c3f2d02>, 검색일 : 2017. 08. 21

김승일 (2014), 기계학습/머신러닝 기초(Machine Learning Basics), <http://www.whydsp.org/237>, 검색일 : 2017.09.19

강미주 (2015), 특집기획 해운분야 ICT 혁명, 새로운 해운시대 열리나 (2015.5.4.), <http://www.monthlymaritimekorea.com/news/articleView.html?idxno=16027>, 검색일 : 2017. 12. 05

조지민 (2017), 한화토탈, 스마트 영업·물류 시스템 오픈...모바일로 실시간 처리·조회, <http://www.fnnews.com/news/201711090833201676>, 검색일 : 2017. 12. 05

임예리 (2017), 자동 라우팅, 물류효율 높일 밑거름 될까, <http://logiseconomy.tistory.com/3322>, 검색일 : 2017. 12. 05

Choi, C. (2014), UPS, Big Data로 배송 적시성 확보하기,

- <https://jacesky1.wordpress.com/2014/10/15/ups-big-data로-배송-적시성-확보하기/>, 검색일 : 2017. 12. 05
- 정유진 (2014), 글로벌 물류기업, 빅데이터 적극 활용, http://www.dt.co.kr/contents.html?article_no=2014031002011132748001, 검색일 : 2017. 12.05
- 최재경 (2016), 빅데이터 분석의 국내외 활용 현황과 시사점, http://www.kistep.re.kr/c3/sub3_2.jsp?, 검색일 : 2017.12.05
- 이영희 (2017), 부산항만공사 해운물류 빅데이터 분석 시스템 구축, <http://www.yonhapnews.co.kr/bulletin/2017/09/14/0200000000AKR20170914152500051.HTML?input=1195m>, 검색일 : 2017.12.05
- 이연경 (2017), 싱가포르항-IBM사, AI기반 해운·항만운영 개선 시스템 시험운영, <http://www.kmi.re.kr/web/trebook/list.do?rbsIdx=135>, 검색일 : 2017.12.05
- KMI 해사안전연구실 facebook (2017), <https://www.facebook.com/kmisafetypolicy/posts/1871285723193043>, 검색일 : 2017.12.05
- Port Technology (2017), Start Up Aims Big Data at Shipping Sector, https://www.porttechnology.org/news/start_up_aims_big_data_at_shipping, 검색일 : 2017.12.05.
- 건국대학교 경영대학원, 의사결정지원시스템의 개요, <https://home.konkuk.ac.kr:8080/cms/Common/MessageBoard/ArticleFile.do?id=53c9c32>, 검색일 : 2017.12.05
- doopedia(<http://www.doopedia.co.kr>), 검색일 : 2017.09.21
- IHS(<https://www.ihs.com>), 검색일 : 2017. 09. 22
- 정해용, 금융 빅데이터 전쟁 - 국내 금융사, 빅데이터 활용 쟁점(2017.06.13.), http://biz.chosun.com/site/data/html_dir/2017/06/12/2017061201897.html, 검색일 : 2017. 12. 5
- 배준호, 스타 머니매니저 시대가 끝나고 있다(2017.03.29.), <http://www.etoday.co.kr/news/section/newsview.php?idxno=1473633#csidxd10cf88eebf0b05858dfe902ebcc28c>, 검색일 : 2017. 12. 5

두산백과, [http://terms.naver.com/entry.nhn?docId=2807097&cid=40942& categoryId=32840](http://terms.naver.com/entry.nhn?docId=2807097&cid=40942&categoryId=32840), 검색일 : 2017.12.05

매일경제용어사전, <http://terms.naver.com/entry.nhn?docId=3360&cid=43659&categoryId=43659>, 검색일 : 2017.12.05.

위키백과, <https://ko.wikipedia.org>, 검색일: 2017.8.21.

국가물류통합정보센터, <https://www.nlic.go.kr>

국세청, <https://www.customs.go.kr>

부록 : 설문조사 《

빅데이터를 활용한 시장분석의 사전작업으로서 해운의사결정, 전통적 시장분석, 빅데이터에 대한 인지도 등에 관한 현황을 파악하기 위하여 설문조사를 실시하였다. 설문조사는 국내 해운기업을 대상으로 진행되었으며 대상기업의 선정에는 업종, 규모가 고려되었다.

1. 표본추출

표본추출에는 자의적 추출 방법이 선택되었다. 계층표본의 방식을 선택하는 것이 바람직하나 한 계층에 속하는 선사의 수가 제한적인 경우가 많아 적용이 곤란하였기 때문이다. 선사를 선택한 후 부분적으로 임원을 대상으로 한 비구조적 면접을 진행하였고 실제 응답자의 선택은 해당 임원이 선정하는 방식을 채택하였다. 외부에서 응답자를 임의로 추출할 경우 응답능력이 없는 대상이 선택될 경우가 많을 것으로 판단하였다.

설문지가 배포된 회사는 총 11개사이고 설문지는 총 68부가 배포되었다. 회사별 설문지의 배포, 회수된 부수는 조직의 규모를 고려하여 적절히 조정하였다. 회수된 설문지는 총 55부로서 응답률은 80.9%이다.

2. 설문지

안녕하십니까? 바쁜 일정에도 설문조사에 응해주셔서 대단히 감사합니다!

한국해양수산개발원은 해양, 수산, 해운, 항만 분야 제 부문의 과제를 체계적으로 연구하고, 이에 관련한 각종 동향과 정보를 신속히 수집 분석 보급함으로써, 국가의 정책수립과 국민경제의 발전에 이바지함을 목적으로 설립된 정부출연 연구기관입니다.

현재 저희 개발원에서는, 빅데이터 분석(big data analytics)*을 통한 해운업의 의사결정 지원을 목적으로, 반복되는 해운업 실패의 예방 및 해운업 경쟁력 제고 방안에 대한 연구를 진행하고 있습니다.

이에, 아래와 같은 내용으로 설문조사를 실시하고자 하오니, 각별한 관심과 함께 많은 응원을 부탁드립니다. 본 설문조사는 국내의 해운 기업을 대상으로 실시하며, 조사결과는 빅데이터 분석 모델 설계와 시스템 구축의 기초 자료로 활용될 것입니다. 귀하의 응답 내용은 저희 연구를 위한 목적으로만 활용될 것이며, 응답자의 익명성은 철저히 보장됩니다.

귀하의 귀중한 실무 경험을 바탕으로 성실하게 응답해 주시면 감사하겠습니다.

2017년9월

연구책임자 윤희성 부연구위원

한국해양수산개발원 해운시장분석센터

* 이 연구에서 정의하는 빅데이터는 대량(volume), 고속 생성(velocity), 다양한 형태(variety)로 정의되는 통상적인 빅데이터에 추가하여 기존 기업 내 외부의 데이터로서 광범위하게 산재하는 전통적 데이터를 포괄 하는 개념입니다.

해운 의사결정에 대하여

1. 데이터를 확보하고 분석하는 궁극적인 목적은 ‘의사결정’에 있습니다. 이에, 해운기업의 의사결정 유형과 그 중요도를 파악하고자 합니다. 아래는 해운업의 의사결정을 대별하여 열거한 내용입니다. 각 항목에 대하여 **중요도 순에 따른 순위**를 매겨 주시기 바랍니다. 만약, 아래 내용 중에 귀사에 있어서 중요한 의사결정이 빠져있다면, 추가 란에 직접 기재해 주신 다음, 순위를 매겨 주십시오(이하, 동일 요령). 아울러 귀하의 의사결정 유형별 근무 기간을 간단히 기재하여 주시면 감사하겠습니다.

구분	순위	기간
• 신규사업진입, 사업확장, 사업축소, 사업철수, 전략적 제휴 등 사업구조의 결정		
• 선박의 확보(선박구매, 용선 등) 및 선박의 판매(수송계약, 대선 등)의 시점 및 계약기간의 선택		
• 거래처의 선정 및 신용의 판단		
• 선속의 증가, 선원비 절감 등 운영효율(투입 대비 산출) 제고 방안의 선택		
• GRI 등 판매 및 구매가격의 결정		
• 추가1 :		
• 추가2 :		
• 추가3 :		

2. 의사결정에는 다양한 준거가 있는 경우가 많습니다. 아래에 열거된 준거 중 실제 귀사에서 **활용도가 높은 순**으로 순위를 매겨 주십시오.

구분	순위
• 정량적, 정성적 데이터분석의 결과로 제시된 내부 분석보고서	
• Marsoft, MSI 등 해운시황예측 전문기관의 예측/전망 서비스	
• Clarkson, BRS, SSY, FIS, Commodore, KMI 등 유료 또는 무료로 확보할 수 있는 분석 보고서	
• SIN(ClarksonShipping Intelligence Network), IHS, Baltic 등이 제공하는 가공되지 않은 데이터(raw data)	
• Lloyd's List, Tradewinds 등 신문, 방송 매체	
• 업계 지인 등 인적 전문가 네트워크로부터 입수되는 시황에 대한 견해(expert opinion)	
• 브로커, 대리점 등 거래처가 전화나 이메일로 제공하는 정보	
• 추가1 :	
• 추가2 :	
• 추가3 :	

해운시장 분석에 대하여

3. 아래에 열거한 시장분석 기법 중, 실제로 귀사에서 **활용도가 높은 순**으로 순위를 매겨 주십시오.

구분	순위
• 시계열 분석(time-series analysis)	
• 기술적(技術的) 분석(technical analysis)	
• 기술(記述)통계분석(descriptive statistics)	
• 델파이법(전문가의 주관적 의견 종합)	
• 시나리오플래닝	
• SWOT분석	
• 추가1 :	
• 추가2 :	
• 추가3 :	

4. 시장분석 조직에 대한 다음의 설명 중, 귀사의 조직에 해당되는 내용을 **모두 선택**해 주십시오(○×식).

• 영업조직과는 별도로 시장분석을 담당하는 조직이 있다.	
- 별도의 시장분석 조직이 연속성을 가지고 운영된다(존폐가 반복되지 않는다)	
- 시장분석 조직의 구성원은 10명 이상이다.	
- 시장분석 조직의 구성원은 5명 이상 10명 미만이다.	
- 시장분석 조직의 구성원은 3명 이상 5명 미만이다.	
- 시장분석 조직의 구성원은 1~2명이다.	
• 영업조직에서 시장분석기능을 담당한다.	
- 영업본부내에 별도의 담당자가 시장분석 기능을 전담한다(영업 기획 등)	
- 영업팀장, 용선 담당자 등이 시장분석을 담당한다.	
• 시장분석 기능이 별도 조직적으로 이루어지지 않고 주로 경영층의 운영 방침 및 지시에 따른다.	

5. 다음은 시장분석 결과의 실제 활용에 대한 서술입니다. 귀사에 해당되는 내용을 **모두 선택**해 주십시오(○×식).

• 시장분석 조직의 의견이 실제 의사결정 과정에서 활용된다.	
- 시장분석 조직의 의견이 의사결정에 잘 반영되는 편이다.	
- 시장분석 조직의 의견이 부분적으로 수용된다.	
- 시장분석 조직의 의견이 거의 받아들여지지 않는 편이다.	
• 시장분석조직의 의견이 활용되는 정도는, 최고 의사결정권자의 리더십 성향에 따라 크게 변화한다.	

6. 다음은 계량적 시장분석에 대한 귀사의 의견을 묻는 내용입니다. 귀사의 의견에 해당되는 항목을 **모두 선택**해 주십시오(○×식).

• 계량적 시장분석은 의사결정에 도움이 되는 유용한 판단기준이다.	
• 계량적 시장분석은 의사결정의 참고자료 중의 하나로 활용되는 정도이다.	
• 계량적 시장분석은 신뢰도가 낮아 의사결정에 거의 도움이 되지 않는다.	
• 계량적 시장분석을 활용한 경험이 없어서 잘 모르겠다.	

활용 데이터에 대하여

7. 아래는 시장분석에 활용되는 데이터의 공급원을 열거한 것입니다. 귀사에서 활용하시는 데이터 공급원의 **활용도** (높음:◎보통:○낮음:△)를 표기해 주십시오. 만약, 귀사가 활용하시는 데이터 유형 및 공급원이 누락되어 있다면 추가 란에 직접 기재하시고 그 활용도를 표기해 주시면 됩니다.

• 전문기관의 예측, 전망보고서	
- MSI	
- Marsoft	
- 추가:	
• Clarkson, Commodore, KMI 등 유료 또는 무료로 확보할 수 있는 시황 분석 보고서	
- Clarkson	
- Commodore	
- Howe Robinson	
- BRS	
- SSY	
- KMI	
- MEIC	
- 추가:	
• 가공되지 않은 데이터(raw data)	
- SIN(Clarkson Shipping Intelligence Network)	

- IHS	
- Baltic Exchange	
- 추가:	
• 뉴스 매체	
- Lloyd's List	
- Tradewinds	
- Hellenic Shipping News	
- AWSJ	
- JOC	
- 한국해운신문	
- 해사신문	
- 쉬핑가제트	
- 쉬퍼스저널	
- 쉬핑데일리	
- 추가 :	
• 추가유형1 :	
- 추가:	
• 추가유형2 :	
- 추가:	
• 추가유형3 :	
- 추가:	

빅데이터 분석에 대하여

8. 아래는 귀사의 빅데이터 이해도에 관련된 질문입니다. 귀사에 해당되는 내용을 **모두 선택**해 주십시오.

• 빅데이터와 전통적 데이터의 차이를 알고 있다.	
• 빅데이터 분석과 그 기법들에 대하여 적어도 개략적인 이해를 하고 있다.	
• 빅데이터 분석과 4차산업혁명의 차이를 설명할 수 있다.	

9. 아래는 빅데이터와 해운업의 관련성에 대한 질문입니다. 귀사에 해당하는 내용을 **모두 선택**해 주십시오.

• 빅데이터가 해운업에 미치는 영향에 대하여 생각해 본 적이 있다 .	
• 빅데이터 분석이 해운 의사결정에 긍정적인 영향을 미칠 것으로 생각한다.	
• 해운업 분야에 있어서 빅데이터 분석은 필요없다고 생각한다 .	

10. 아래는 귀사의 빅데이터 활용에 대한 질문입니다. 해당되는 내용을 **모두 선택**하시고, 그 내용에 대해 기술해 주십시오.

• 우리 회사는 빅데이터를 수집하고 있다.	
- 수집항목 1 :	
- 수집항목 2 :	
- 수집항목 3 :	
• 우리 회사는 빅데이터를 분석하고 있다.	
- 분석기법 1 :	
- 분석기법 2 :	
- 분석기법 3 :	
• 우리회사는 빅데이터 분석의 결과를 의사결정에 활용하고 있다.	
- 의사결정영역 1 :	
- 의사결정영역2 :	
- 의사결정영역3 :	

응답자 정보

응답자에 대한 정보를 여쭙습니다. 귀사의 회사명은 회사별 응답 숫자 파악에만 활용되며, 아래 항목 중의 어떤 내용도 응답자를 식별하는 데 사용되지 않습니다.

• 회사명	
• 회사의 업종	선주업() 컨테이너() 벌크() 종합해운() 그 외()
• 담당직무분야	기획() 용대선() 마케팅() 그 외()
• 직위	임원() 팀장() 담당() 그 외()
• 해운업 종사기간	()년
• 의사결정 업무 담당 기간	()년

설문조사에 협력해 주셔서 대단히 감사합니다!

빅데이터를 활용한 해운시장분석 연구 - 빅데이터의 파악 및 확보방안

• 인 쇄	2017년 12월 29일 인쇄
• 발 행	2017년 12월 31일 발행
• 발 행 인	양 창 호
• 발 행 처	한국해양수산개발원 49111 부산시 영도구 해양로 301번길 26(동삼동)
• 연 락 처	051-797-4800 (FAX 051-797-4810)
• 등 록	1984년 8월 6일 제313-1984-1호
• 조판·인쇄	(주)디자인월드 Tel : 051-916-1533

판매 및 보급 : 정부간행물판매센터 Tel : 394 - 0337

정가 6,000원