http://www.calsec.or.kr/jsebs
ISSN: 2288-3908

The Journal of Society for e-Business Studies
Vol.19, No.3, Aug 2014, pp.51-64
http://dx.doi.org/10.7838/jsebs.2014.19.3.051

철강산업 트렌드 분석을 위한 텍스트 마이닝 도입 연구 : P社 사례를 중심으로

A Pilot Study on Applying Text Mining Tools to Analyzing Steel Industry Trends : A Case Study of the Steel Industry for the Company "P"

민기영(Ki Young Min)*, 김훈태(Hoon Tae Kim)**, 지용구(Yong Gu Ji)***

초 록

기업은 생존을 위해 수많은 정보 속에서 빠르게 상황을 인식하고 미래를 예측하기 위해 정량데이터 분석뿐만 아니라 비정형데이터 분석에 대한 관심이 높아지고 있으나, 철강산업에서는 아직 활발하게 활용되지는 못하고 있다. 이에 본 연구에서는 글로벌 철강회사인 P社의 사례를 중심으로 텍스트 마이닝을 이용한 산업트렌드 분석을 시도해 경쟁사 전략, 관심국가의 시장변화, 해외사업장 여론 등을 파악 하는데 기여할 수 있다는 가능성을 발견하였다. 사례분석은 철강산업을 10개의 카테고리로 분류하고 각각 10개의 주제를 선정하여 분석을 시도하고, 이중 의미 있는 변화를 발견하면 심층 분석하는 형태로 진행하였다. 이번 P社의 사례를 통해 텍스트 마이닝을 통한 산업트렌드 분석이 더 의미 있기 위해서는 목적을 명확히 하고, 관련 키워드를 체계화한다면 경쟁사 전략 파악, 리스크관리, 정량데이터 예측 보정 등 많은 부분에 기여할 수 있을 것으로 기대한다.

ABSTRACT

It becomes more and more important for business survival to have the ability to predict the future with uncertainties increasing faster and faster. To predict the future, text mining tools are one of the main candidate other than traditional quantitative analyses, but those efforts are still at their infancy. This paper is to introduce one of those efforts using the case of company "P" in the steel industry. Even with only four month pilot studies, we found strong possibilities, if not testified robustly, to predict future industrial trends using text mining tools. For these text mining case studies, we categorized steel industry trend keywords into ten components (10 categories) to study ten different subjects for each category. Once found any meaningful changes in a trend, we had investigated in more detail what and how some trend happened so. To be more roust, firstly we need to define more cleary the purpose of text mining analyses. Then we need to categorize industry trend key words in a more systematic way using systems thinking models. With these improvements, we are quite sure that applying text mining tools to analyzing industry trends will contribute to predicting the future industry trends as well as to identifying the unseen trends otherwise.

키워드 : 텍스트 마이닝, 트렌드 분석, 철강산업
Text Mining, Trend Analysis, The Steel Industry

* First Author, Graduate Program in Technology Policy, Yonsei University, Center for Economy Research and Analysis, Posco Research Institute(mky0919@posri.re.kr)
** Co-Author, Graduate Program in Technology Policy, Yonsei University, Center for Economy Research and Analysis, Posco Research Institute(htkim77@posri.re.kr),
*** Corresponding Author, Information and Industrial Engineering, Yonsei University(yongguji@yonsei.ac.kr)
2014년 06월 18일 접수, 2014년 08월 11일 심사완료 후 2014년 08월 14일 게재확정.

1. 서 론

기업이 산업의 트렌드를 분석하고 경영환경을 예측하기 위해 지금까지는 주로 정량데이터 분석이나 설문조사 기법을 이용해 왔다. 그러나 시장 환경이 복잡해지고 불확실성이 커지면서 제한된 지표와 함수로는 시장의 전반적인 메커니즘을 이해하고 예측하기에는 정확도가 낮아지고 있는 추세이다[14].

이에 많은 기업들이 비정형데이터 분석 기법인 '텍스트 마이닝'에 관심을 보이고 있다. 텍스트 마이닝은 실시간으로 생산되는 수많은 비정형데이터 속에서 의미 있는 패턴을 발견해 고객의 마음을 읽거나 시장을 예측하는데 활용되고 있어, IT산업이나 패션산업 등 소비자의 접점에 있으면서 변화가 빠른 산업분야에서 마케팅과 산업트렌드 분석에 활용되고 있다[7].

반면 철강, 석유화학 등 설비 중심의 장치산업은 주문생산체계여서 상대적으로 외부 변화에 덜 민감한 편이였지만, 2000년대 초부터 중국의 산업발전과 글로벌 경기침체 장기화로 외부의 변화에 관심이 높아지고 있다[11].

이러한 맥락에서 글로벌 철강회사인 P社도 글로벌 네트워크에서 수집해서 보내오는 수많은 비정형데이터를 관리하는 'Global Information Hub System'을 운영해 중요한 정보를 찾아 의사결정자에게 전달하는 등 매일매일의 정보흐름을 파악하는 역할을 수행해 왔다. 그러나 대부분의 철강산업 트렌드는 정량데이터 분석 모델을 통해 예측하다 보니 시시각각으로 변화하는 경쟁사나 관심국가의 트렌드를 읽어내기에는 한계가 있어, 정량데이터 분석으로 발견할 수 없는 부분을 텍스트 마이닝을 통한 비정형데이터 분석을 통해 보완할 수 있는지 연구해 보고자 한다.

이에 본 연구에서는 P社의 'Global Information Hub System'을 운영하면서 축적된 데이터베이스와 정량데이터 분석 모델에서 사용된 지표를 참고해 철강산업 트렌드 분석을 위한 관심 분야를 도출하고, 관심 분야의 주제별 키워드를 중심으로 비정형데이터를 수집해 텍스트 마이닝을 통해 철강산업의 트렌드를 파악하는데 도움이 되는지, 그리고 정량데이터 분석을 통해 파악할 수 없었던 부분을 발견하는데 도움을 줄 수 있는지를 찾아보는데 목적이 있다. 또한, 본 연구를 통해 텍스트 마이닝으로 산업의 트렌드를 더 잘 분석하기 위해 남은 과제는 무엇인지 제시해 산업의 트렌드 분석에 관심 있는 많은 분들의 추가적인 의견을 구하고자 한다.

본 연구의 구성은 다음과 같다. 제 2장에서는 트렌드 분석과 텍스트 마이닝에 대한 이론적 배경을 제시하였고, 제 3장에서 연구방법을 제시하고, 제 4장에서는 철강산업의 트렌드 분석결과와 텍스트 마이닝 도입 사례를 소개하고, 제 5장에서는 결론과 논문의 한계와 향후 과제를 제시하였다.

2. 이론적 배경

2.1 산업 트렌드 분석의 흐름

트렌드란 뭔가가 현재 '진행 중'이거나 어떤 방향을 향해 가고 있는 경향이나 추세, 동향, 유행 등을 말하는 것으로 트렌드는 시간의 변화와 관련이 깊고 과거와 현재의 데이터들을 모아 미래에 대한 가설을 만드는 것이라고 할 수 있다[14].

트렌드 분석의 목적은 과거의 트렌드를 통해 현재의 트렌드가 무엇인가를 밝혀내는 현상파악(Fact)과 이렇게 파악된 현상을 토대로 미래를 전망하고 예측하기 위한 것이다[14].

트렌드 분석의 방법은 크게 귀납적 접근법과 연역적 접근법으로 나눌 수 있다. 귀납적 접근법은 사회과학이나 심리학 등에서 많이 사용하는 방법으로 다양한 케이스들을 지속적으로 살펴서 예측하는 방법으로 결론을 내기 어렵고 주관적인 경향이 있고, 연역적 접근법은 경제학이나 시장을 분석하는 데 많이 사용하는 방법으로 분석 이론이나 모델을 통해 예측하는 방법으로 이론이나 모델이 상당히 발달되어 있어 과학적 접근이 가능하다[8].

연역적 접근 방법의 기본은 이론이나 모델에 의해 핵심지표를 선정하고 핵심지표 간의 상관관계와 인과관계를 분석하는 것인데, 이러한 방법의 한계는 모델에 사용한 핵심지표 이외에 '모든 지표는 변하지 않는다는 것(Ceteris Paribus)'을 전제로 설계되어, 세상의 복잡성이 커지면서 분석에서 제외된 변수들의 영향이 점점 커지고 있어 예측력은 떨어지고 있다[8]. 이러한 문제를 해결하기 위해 정태적 선형예측모델은 시간과 확률함수를 이용해 동태적 예측모델로 발전해 나가고 있으며, 지표간의 관계 분석은 연립방정식이나 산업연관분석, 시스템 다이나믹스 기법 등을 이용해 시스템 전체의 상관관계를 살피려는 노력을 하고 있다.

그러나 이러한 노력에도 불구하고 산업연관분석은 5년 단위 분석으로 수시로 변하는 세상의 변화를 수용하기 어려운 측면이 있고, 시스템 다이나믹스 기법은 전체의 흐름을 계량화하고 상관관계를 파악하기가 어려워, 텍스트 마이닝을 통해 이러한 애로사항을 보완하려는 시도가 진행되고 있다.

2.2 전통적인 철강산업 트렌드 분석의 한계

철강산업은 지금까지 변화의 동인이 적은 편이라 수급이나 가격 등 시장 분석은 정태적 선형예측모델이나 단순회귀분석을 많이 사용해 왔으며, 대표적인 분석방법으로는 개별지표에 의한 방법, 종합경기(선행)지수에 의한 방법, 계량모형에 의한 방법 등을 들 수 있다[11].

첫 번째, 개별지표에 의한 방법은 철강경기의 움직임을 가장 잘 반영한다고 판단되는 개별 지표들의 추이를 분석해서 종합적으로 판단하는 방법으로 특정 지표에 의존하기 때문에 전체 철강산업의 움직임을 포괄적으로 파악하기에는 어려움이 있고 분석자의 주관에 치우치기 쉬운 단점이 있다. 두 번째, 종합경기(선행)지수에 의한 방법은 산업의 움직임을 잘 반영하는 개별 지표들을 선정한 후 이를 통계적으로 가공해 종합 경기지수를 산출해 이것이 과거의 움직임으로부터 유의적인 규칙성이나 유사성을 발견해서 현재의 경기가 어느 시점과 유사한지를 보고 판단하고 예측하는 방법이다. 세 번째, 계량모형에 의한 방법은 변수들 간의 인과관계나 상관관계를 계량모형화하고 이를 이용하여 경기를 예측하는 방법으로, 철강산업의 여건이나 구조가 바뀌면 변수들 간의 관계도 변해 모형의 설득력이 저하되고 일반인이 쉽게 이해하기 어려운 단점이 있어 주로 종합경기지수 방법과 계량모형을 혼용한 형태를 많이 사용해 왔다. 이러한 방법들은 현상은 잘 설명하지만 모델에 사용한 핵심지표 이외에 '모든 지표는 변하지

않는다는 것(Ceteris Paribus)'를 전제로 하기 때문에 예측력은 떨어지고 있다.

철강산업은 2000년대부터 중국이 철강수입국가에서 수출국가로 부상하면서 원료가격이 상승하고 생산량은 증가하면서 공급과잉이 심각해졌고, 철강기술의 발전은 주문생산체계였던 철강제품의 상품화(Commoditization) 즉 범용(Commodity)으로 변화되고 있다는 것이다[11]. 이에 단편적인 시장예측에서 벗어나 경쟁사 전략이나 관심국가의 정책변화를 선제적으로 파악해 대응하는 것이 중요하게 되어 전통적인 트렌드 분석 방법을 보완할 새로운 방법론의 연구가 필요하게 되었다.

2.3 텍스트 마이닝

텍스트 마이닝(Text Mining)이란 대규모의 문서(Text)에서 의미 있는 정보를 추출하는 것을 말하며, 분석 대상이 비구조적인 문서정보라는 점에서 데이터 마이닝과 차이가 있다[7].

정량데이터를 다루는 통계학과 비정형데이터 다루는 텍스트 마이닝의 차이를 살펴보면 통계학은 비교적 크지 않은 실험데이터를 대상으로 하는데 반해 텍스트 마이닝은 비계획적으로 축적된 대용량의 데이터를 대상으로 한다[6]. 통계학이 추정(Estimation)과 검정(Testing)이라는 이론을 중시하는 특징을 가졌다면 텍스트 마이닝은 기업 활동 과정에서 자연스럽게 축적된 대량의 데이터를 분석해 기업 경영에 필요한 가치 있는 정보를 추출하기 위해 사용된다. 이러한 이유로 텍스트 마이닝을 “규모, 속도, 그리고 단순성의 통계학(Statistics at Scale, Speed, and Simplicity)”이라 부른다[4].

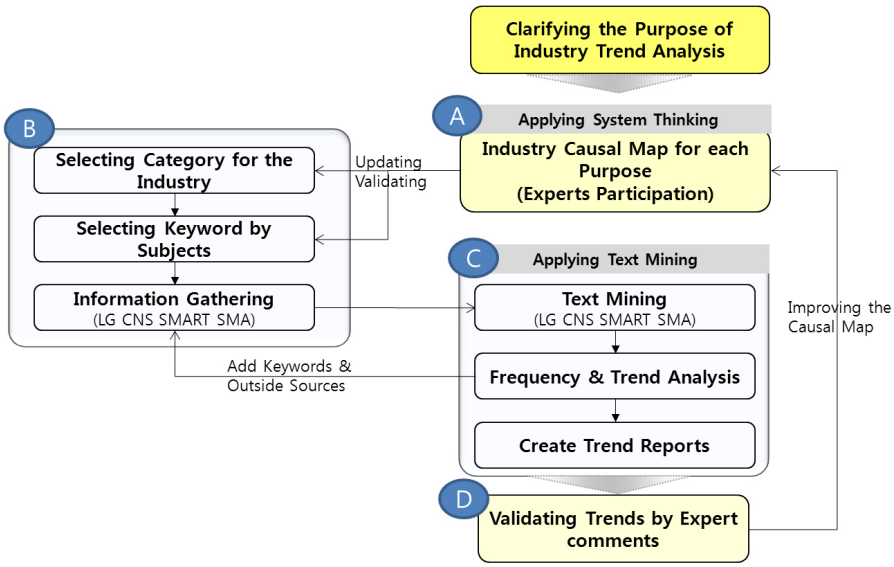
텍스트 마이닝은 데이터 준비 단계(Text Refining)와 지식 추출(Knowledge Distillation) 과정으로 구분할 수 있으며, 데이터 준비 단계는 다양한 정보원(Information Source : 인터넷, 인트라넷, 트윗터 등)에서 수집한 자유로운 형태의 텍스트 문서를 중간 형태(Intermediate Form)로 바꾸는 단계로 자연어처리, 웹문서의 경우 태그 제거, URL과 타이틀 추출 등의 기술체계를 포함하고 있다[2]. 지식추출 과정은 중간 형태의 문서에서 의미 있는 패턴과 지식을 유추해 내는 과정으로 클러스터링, 분류, 시각화, 문서요약, 기계학습 등의 기술체계를 포함하고 있다[2].

텍스트 마이닝 결과를 시각화하는 방법 중 하나인 단어 구름(Word Cloud)은 문서에 사용된 단어의 빈도를 계산해서 시각적으로 표현하는 것으로 많이 나오는 단어는 크게 표시되기 때문에 한 눈에 문서의 핵심 내용을 파악할 수 있다[7].

3. 연구 방법

지금까지 철강산업 트렌드는 수요와 공급을 중심으로 하는 부분균형분석(Partial Equilibrium Analysis)을 실시해 왔다. 그러다보니 경쟁사 전략이나 통상이슈 등 많은 변수들이 존재하지만 이를 파악하기도 어렵고, 정량화하기도 어려워 불변이라고 가정하고 분석해 왔다.

이에 본 연구에서는 지금까지 철강산업 트렌드 분석에 사용되지 않았던 변수나 참고요인들을 비정형데이터 분석을 통해 찾아내기 위해 <Figure 1>과 같은 절차로 분석을 진행하였다.



〈Figure 1〉 Steel Industry Trend Analysis Applying Text Mining Tools

먼저 철강산업 트렌드 분석을 위해 산업 전반의 흐름과 상호작용을 파악할 수 있도록 P社에서 운영하던 'Global Information Hub System'에 축적된 데이터베이스에서 분야별 키워드를 추출하고[12], 분야별 전망 예측모델에서 활용되는 변수를 참고, 시스템적 사고(System Thinking)를 통해 철강산업 인과지도를 도출하였다[9](<Figure 1>의 Ⓐ).

철강산업 인과지도는 철강산업을 전체적으로 모니터링 할 수 있도록 10개의 카테고리로 나누고, 각 카테고리별로 인과지도를 만들어 10개의 주제를 도출하고 이를 대표하는 핵심 키워드를 선정하여 이를 기반으로 정보를 수집하였다(<Figure 1>의 Ⓑ).

이렇게 수집된 정보를 가지고 'LG CNS의 SMART SMC'를 활용 텍스트 마이닝을 적용해 제공하는 분석결과 중 의미 있는 트렌드를 발견해 트렌드 분석 리포트를 만들어(<Figure 1>의 Ⓒ), 전문가의 평가를 거쳐 트렌드로 채택 여부를 결정하였다(<Figure 1>의 Ⓓ).

3.1 관심 키워드 정의 및 데이터 수집

체계적으로 정보를 분석하기 위해 무엇보다 중요한 것은 가치 있는 정보를 수집하기 위한 포커싱된 탐색 작업이다. 정보의 수집은 철강산업 트렌드 파악을 위한 미시영역과 거시영역을 포괄해 철강제품, 철강시황, 철강기술, 원료, 소재, 수요산업, 경쟁사, 관심국가, 철강정책 등 10개의 카테고리로 나누고, <Table 1>과 같이 그 동안 축적된 'Global Information Hub System'의 데이터베이스를 바탕으로 카테고리별 주제를 10개 선정하고 주제별 키워드를 선정하였다. 각 주제별 키워드를 중심으로 국내외 주요 언론사, 철강관련 기관이나 협회, 경쟁사 홈페이지, 트위터 등 200여 개(한글, 영어)의 사이트를 대상으로 2014년 2월부터 5월까지 4개월 동안 트렌드 분석을 실시하였다.

〈Table 1〉 Subjects by 10 Categories for Steel Industry Trend Analysis

Category	Subjects
Steel Products	Stainless steel, Pipe, Shape, Plate, Cold rolled, Hot rolled, Galvanized steel sheet, etc
Steel Market	Production, Price, Export, Import, Demand, Distribution, Inventory, etc
Steel Technologies	Weight reduction, Rolling, Plating, Steel making, Ironmaking, Finex, etc
Raw Materials	Nickel, Iron ore, Coaking coal, Scrap, etc
Non-Metallic Materials	Aluminum, lithium, Magnesium, Titanium, Petrochemistry, Carbon fiber, Rare Earth, etc
Competing Steel Makers	US Steel, NUCOR, NSSMC, Baogang, Arcelor Mittal, etc
Steel Demand Industries	Automotive, Construction, Electronics, Shipbuilding, Energy, Machinery, Civil engineering, etc
Company “P”	Investment, Prospect, Competitiveness, Stock price, Performance, Finance, Social contribution, etc
Major Countries	China, USA, Japan, Europe, Indonesia, Russia, etc
Steel Related Policies	Fair trade, Waste, FTA, Certified Emission Reductions, Power charges, Protectionism, etc

3.2 텍스트 마이닝 적용 및 트렌드 발굴

본 연구에서는 텍스트 마이닝 결과의 질을 높이기 위한 방법보다는 텍스트 마이닝 결과를 통한 트렌드 발굴에 주력하였다. 텍스트 마이닝 툴은 LG CNS의 소셜미디어 분석 솔루션 ‘SMART SMA’를 활용하였으며[13], 텍스트 마이닝 분석은 ‘SMART SMA’에 100개의 관심 주제별 키워드와 대상 사이트를 지정해 놓으면 ‘SMART SMA’가 대상 사이트에서 자동으로 데이터를 수집해 텍스트 마이닝 분석결과를 제공한다. 이렇게 제공된 분석결과를 가지고 주제별 키워드의 빈도수나 추이[1], 단어 간의 상관관계 등을 살펴, 평상시와 다른 변화의 징후가 발견되는지를 매일 모니터링 하였다. 그 중 새로운 키워드가 등장하거나 빈도수가 급증하는 등 변화의 징후가 보이면 그 원인을 찾기 위해 분석결과를 ‘단어 구름(Word Cloud)’로 시각화해 보고, 추가적인 분석이 필요한 경우에는 분석 목적을 더 구체화하여 대상 키워드와 사이트, 기간을 재설정해서 다시 텍스트 마이닝을 시도해 단어들 간의 상관관계(co-word analysis)를 분석해 의미 있는 패턴을 파악하는 형태로 진행하였다[5, 10]. 이런 분석의 과정을 거쳐 매주 하나씩 가장 의미 있는 아이템을 선정해 트렌드 리포트로 작성해 전문가에게 검증받는 형태로 진행하였다.

3.3 전문가를 통한 트렌드 검증

전문가 검증은 철강산업 전문가 10명(철강업 10년 이상 종사자)의 서면인터뷰를 통해 실시하였다. 전문가들은 텍스트 마이닝을 통해 분석된 트렌드 분석 리포트를 보고, 임팩트 매트릭스(Impact Predictablility Matrix)를 활용해 객관적으로 평가하고 검토의견을 제시하였다. 임팩트 매트릭스(Impact Predictablility Matrix)는 발생 가능성과 비즈니스에 대한 영향을 양축으로 높고 낮음으로 구분한 후 경우의 수에 따라, ‘발생가능성 낮고, 비즈니스 영향 정

도 낮음 = 1점', '발생가능성 높고, 비즈니스 영향 정도 낮음 = 2점', '발생가능성 높고, 비즈니스 영향 정도 낮음 = 2점', '발생가능성 높고, 비즈니스 영향 정도 높음 = 3' 중에 하나를 선택하게 하고, 의견을 제시하도록 하였다[14].

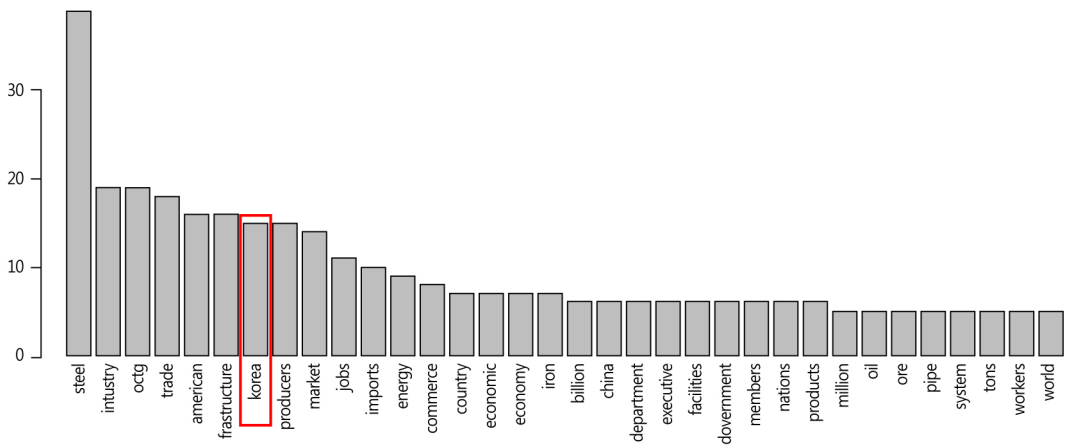
전문가 검증은 평가결과를 합산하여 산술평균을 낸 뒤 평균값에 따라 트렌드의 유효성을 가늠하였는데, 평균값이 2점 이상이면 발생가능성과 비즈니스에 영향도가 높다고 보고 유의미한 트렌드로 채택하고 평균값이 2점 이하면 기각하였다.

4. 텍스트 마이닝 도입 사례

2014년 2월부터 5월까지 4개월 동안 100개의 관심 주제를 중심으로 트렌드 분석을 시도한 결과, 철강산업의 특성상 텍스트 마이닝을 통한 데이터의 흐름이 많이 변하지 않아 새로운 트렌드를 발견하기 쉽지 않았다. 매주 하나씩 발굴한 16개 아이템 중 평균 2점 이상을 받은 아이템은 "미국의 반덤핑 관련 이슈", "중국 철강산업 구조조정 관련 정책변화 예상", "P社 해외 사업장 여론 모니터링 가능성" 등 3개로 나타났고, 나머지 13개는 발생가능성도 비즈니스 영향도 낮아 트렌드로 참고할 만한 가치가 적은 것으로 나타났다. 이는 전체 분석 아이템 중 18.7%가 채택된 것으로 처음 시도된 새로운 방식의 트렌드 분석이라는 측면에서 유용한 결과라고 할 수 있다. 이에 전문가 검증결과 유의미했던 3개의 아이템 중 2개에 대한 분석과정과 분석결과 그리고 전문가 의견을 소개하고자 한다.

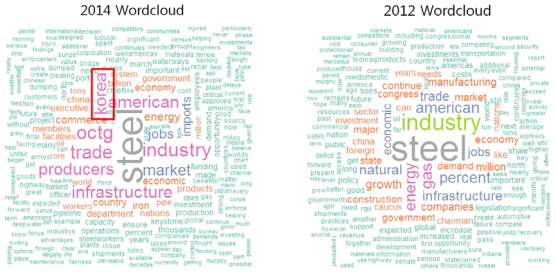
4.1 (사례 1) 美 철강업계, 한국産 반덤핑 무혐의에 불만 고조

경쟁사 영역을 모니터링 하던 중 발견된 사례로, 2014년 3월 25일 미국철강협회(AISI) 주최로 미국의 대표적 철강기업 CEO로 부터 철강업계의 현안과 애로사항을 청취하는 회의(Testimony) 개최한 후, 공개된 회의결과를 분석해 보면 <Figure 2>에서 보는 것처럼 US스틸 등 9개 철강업체에서 "한국"에 대한 키워드가 많이 출현하는 것을 볼 수 있었다[3].



〈Figure 2〉 Frequency Analysis of 9 Steel Company CEO Speeches after AISI Testimony

〈Figure 3〉 Word Cloud Analysis of US Steel, Nucor, and EVRAZ CEO Speeches after AISI Testimony



〈Figure 4〉 Comparison of Word Cloud Analysis between 2012 and 2014 after Respective AISI Tesimony

회사별 발언 정도를 살펴보면, <Figure 3>에서 보는 것처럼 US Steel과 Nucor가 "한국"을 자주 언급한 것으로 나타났으며, EVRAZ는 "한국"이 이번 Testimony의 메인 주제로 중앙에 부각된 것을 알 수 있었다.

또한, <Figure 4>에서 보는 것처럼 지난 2012년과 2014년의 "한국"에 대한 언급 정도를 비교해 봐도 2014년에 그 빈도가 크게 늘어났음을 알 수 있었다.

그 이유는 2013년 7월 한국 등이 미국에 수출한 유정용 강관으로 인해 피해를 봤다며 반덤핑 조사를 청원했으나, 미 상무부는 한국 업체에 대해 부정 판정(Negative Determination)을 내렸다. 그러나 미국 철강업체는 2013년 한국이 미국에 수출한 유정용강관이 894천 톤으로 반덤핑 판정을 받은 나머지 8개국의 수출량을 모두 합친 것보다 많기 때문에 한국의 물량을 줄이지 않는 한 미국 업체의 어려움은 지속될 것이기 때문에 2014년 8월 ITC가 최종 판결을 내릴 때까지 지속적으로 한국의 반덤핑 판정을 요구를 할 것으로 예측되었다. 이에 전략적인 대응에 참고가 필요하다는 트렌드 분석결과를 제시하였다.

이와 같은 트렌드 분석결과에 대해 전문가들은 선제적으로 파악하기 어려운 문제를 텍스트 마이닝을 통해 분석한 상당히 의미 있는 시도라는 의견과 함께 실제적인 대응 준비가 필요하다는 의견을 제시하였다. 전문가 의견을 몇 가지 소개하면 다음과 같다.

"시기적으로 매우 적절한 것 같습니다. 반덤핑의 경우 사전대응이 가장 중요한데 이러한 트렌드 분석을 통해 제소자들 특히, 가장

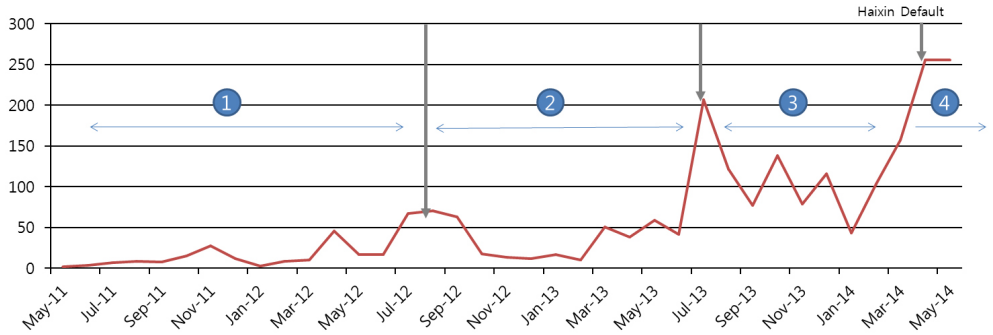
많은 제소를 하고 있는 미국의 철강업자들이 한국산제품 수입에 대해 어떻게 생각하고 있는지를 미리 보여 주는 것이라 할 수 있습니다.", "아주 중요한 이슈를 트렌드 분석을 통해 제시해 주신 것 같습니다. 관련부서에 계신 분들과 함께 이러한 이슈에 대해 추가적인 토론이 이루어지면서 트렌드분석을 다듬어 나간다면 선제적으로 통상리스크를 예방하는 전략을 고민해 낼 수 있을 것이라고 생각됩니다.", "세계 철강경기의 회복이 불투명하고 공급과잉 물량의 경쟁적 수출로 각 수입국들이 반덤핑 등 수입제한 조치를 강화하고 있는 것 같습니다. 반덤핑의 경우 조사 후에 무혐의 판정을 받아내기가 매우 어려우므로 사전적으로 대응하는 것이 매우 중요합니다. 이러한 차원에서 트렌드 분석은 사전적 조치 혹은 준비를 하는데 매우 유용한 정보가 될 것 같습니다."

4.2 (사례2) 바이두(Baidu) 통해서 본 中 철강산업 구조조정 예상

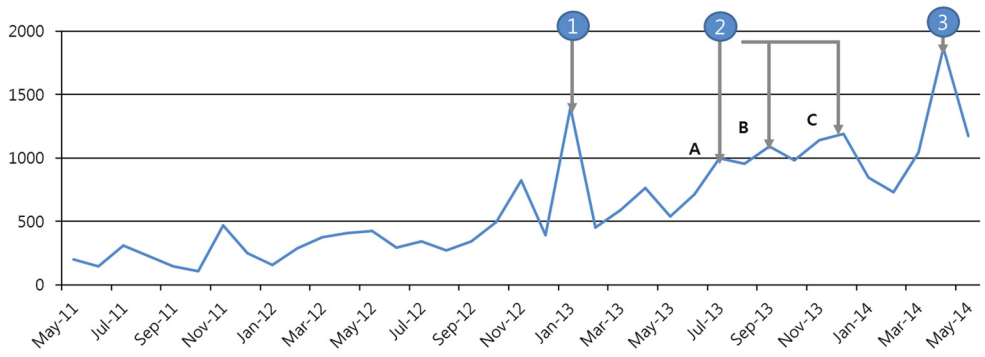
관심지역 중 하나인 중국을 모니터링 하던 중 발견된 사례로, 2014년 3월 31일 '하이신(Haixin)'이라는 철강회사의 부도로 관련 키워드의 빈도수가 크게 증가하고, 그 외에도 '철강산업의 구조조정 가능성'이 자주 거론되고 있어, 둘 간의 상관관계를 파악하기 위해 중국어 검색엔진 바이두(Baidu)에서 2011년 5월부터 2014년 5월까지 3년 간의 뉴스를 대상으로 "철강산업 & 구조조정(Restructuring)", "철강산업 & 부도(Default)"라는 키워드로 검색해 월별 키워드의 발견 추이와 단어들 간의 상관관계(co-word analysis)를 살펴보았다[5].

먼저 <Figure 5>에서 보는 것처럼 "철강산업 & 부도"에 대한 추이는 3개의 봉(峰)을 기준으로 구분이 가능하고 주요매체의 기사 내용을 분석해 보면, ①구간은 '철강산업은 큰 문제없을 것'으로 언급되었고, ②구간은 '중소유통상의 부도 가능성' 정도가 언급되었으며, ③구간에서는 '철강사 부도 가능성'이 언급되었다. ④구간에서는 '철강사는 물론 수요산업까지 부도 가능성'이 언급되고 있었다.

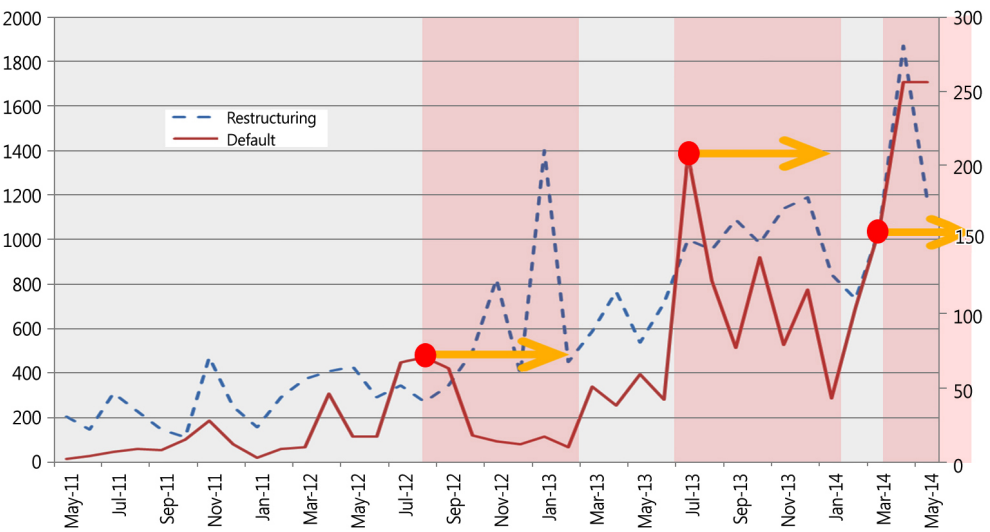
두 번째로 <Figure 6>에서 보는 것처럼 "철강산업 & 구조조정"에 대한 추이를 살펴보면, 2011년에는 200~300건 수준에 불과했으나, 최근에는 1,200~1,800건으로 6배 수준



〈Figure 5〉 Frequency Analysis of a Keyword "Steel and Default"



〈Figure 6〉 Frequency Analysis of a Keyword "Steel and Restructuring"



〈Figure 7〉 Comparisons of Two Frequency Analyses of "Steel and Restructuring" and of "Steel and Default"

으로 폭증하고 있음을 알 수 있었다.

<Figure 6>의 주요 이벤트의 원인을 살펴보면, 피크 ①은 2013년 1월 중국 정부가 산업 인수 및 합병 가이드라인을 발표하였고, 피크 ②의 A는 상반기 철강산업 실적이 발표되었고, B는 구조조정 관련 토론회가 개최되었으며, C는 일부 철강사 구조조정이 진행되었다. 피크 ③은 보산강철 6백만 톤 구조조정 계획 발표되면서, 철강산업이 더욱 악화되어 구조적 변화가 필요 한다는 의견이 많아지는 등 정부정책 혹은 기업의 구조조정 계획 발표에 따라 증가하고 있었다.

마지막으로 <Figure 7>에서 보는 봐와 같이 철강산업의 "부도와 구조조정" 추이를 겹쳐 보면, 부도(실선)가 피크를 치면 구조조정(점선)도 같이 상승하는 것을 볼 수 있었다[10]. 이는 부도와 구조조정 간의 뚜렷한 상관관계를 보이고 있음을 알 수 있었다.

따라서 향후에도 이런 특성이 지속된다고 가정하면,『불황 지속 → 특정 기업 부도 → 구조조정 대책 혹은 계획 발표』의 패턴으로 나타날 것으로 예상된다. 이런 현상은 중국 정부나 기업이 자발적이고 근본적 차원에서 구조조정 계획을 발표하기 보다는 부도 등 이벤트가 발생하면 이에 대한 대책 차원에서 구조조정 계획을 발표하기 때문에 이와 같은 패턴을 보이는 것으로 추정되므로 관련 전문가의 추가적인 분석과 모니터링이 요구된다는 트렌드 분석결과를 제시하였다.

이와 같은 트렌드 분석결과에 대해 전문가들은 정량 데이터로 파악하기 어려운 문제를 텍스트 마이닝을 통해 분석한 매우 흥미로운 시도라는 의견과 함께 추가적인 모니터링과 분석이 필요하다는 의견을 제시하였다. 전문가 의견을 몇 가지 소개하면 다음과 같다.

"매우 흥미로운 분석입니다. 과거 저금리 시절 차입을 통한 설비확장을 했다가 최근 정부가 긴축정책을 실시하면서 자금줄을 조이고, 철강시황은 악화되다보니 일부 기업에서 부도가 나기 시작한 것으로 판단됩니다. 문제는 이러한 상황이 일시적일 것인지 구조적일 것인지를 잘 모니터링하고 분석할 필요가 있다고 보입니다.", "국가 정책이나 일반적 사회이슈 등은 정량적 지표가 없어서 심도 깊은 연구에 어려움이 많았는데, 새로운 기법을 활용해 중국 구조조정과 부도를 활용한 점이 매우 의미 있어 보입니다. 그리고 이를 통해 "불황 지속 → 특정 기업 부도 → 구조조정 대책 혹은 계획 발표"이라는 패턴을 찾아낸 것도 유의미한 발견이라고 생각됩니다. 한 가지 아쉬운 점은 "특정 기업 부도"와 "구조조정 대책 혹은 계획" 사이의 인과관계를 조금 더 명확히 살펴봐야 할 것 같습니다. 향후 충분한 시계열이 확보된다면, 인과관계 검증 등을 사용해 보면 좋을 것 같습니다.", "매우 시의 적절한 분석인 것 같습니다. 정부의 정책의지가 매우 강한 것이 아닌가 여겨집니다. 등량도태 정책이 오래 전에 나왔으나 증가하는 철강수요 때문에 그 효과가 잘 나타나지 않았는데, 중국도 과잉 공급에 직면하고 가격이 저수준에서 지속되다보니 금융지원만으로 위기를 극복하기에는 한계에 도달한 것 아닌가 생각됩니다. 이러한 분석에 더하여 실제로 구조조정이 얼마나 진행되는지도 파악할 수 있었으면 좋겠습니다."

5. 결 론

본 연구는 텍스트 마이닝 도입이 철강산업 트렌드를 분석하는데 도움이 되는지를 P社의 사례를 통해 살펴보았다. 이번 사례 연구를 통해 지금까지 정량데이터 분석으로는 발견할 수 없었던 통상이슈에 대한 리스크를 선제적으로 발견할 수 있다는 가능성을 발견하였고, 경쟁국가인 중국의 시장 변화에 따른 정책 변화 및 기업 구조조정 계획 발표 등 시장의 변화를 조기에 감지할 수 있었다. 그리고 언론보도에 많이 노출되지 않아 확인이 어려운 해외진출 지역의 여론을 모니터링 할 수 있는 방법이 될 수 있다는 가능성을 발견하였다.

본 연구의 결과를 정리하면 다음과 같다.

첫째, 경쟁사의 전략 트렌드를 읽어 내는데 도움이 될 수 있다는 가능성을 발견하였다.

여기서 '도움이 된다'고 단정적으로 결론을 내릴 수 없는 이유는 경쟁사의 트렌드를 분석하기 위한 체계적인 키워드 선정의 경험치가 부족하고, 정보 수집 양도 부족했기 때문이다. 하지만 앞으로 이런 부분을 좀 더 보완하고 지속적으로 데이터를 쌓아 간다면 더 의미 있는 결과를 도출할 수 있을 것이라고 확신한다. 또한, 이렇게 비정형데이터 분석을 잘 발전시켜 간다면 어느 기업이든 경쟁사 전략을 알아내기 위해 사용되던 비도덕적인 산업스파이 활동이 아니라 공개된 정보 분석을 통해 CIA(미국의 중앙정보부) 등과 같은 정보경쟁력을 가질 수 있는 것이다.

둘째, 정량데이터 분석 중심의 수급이나 가격 예측을 보정하는 데도 일정부분 기여할 수 있을 것이라 생각한다. 지금까지 해오던 정량데이터 분석을 통한 수급과 가격 예측에 추가적으로 수급과 가격에 영향을 미칠 수 있는 키워드를 체계적으로 도출해 텍스트 마이닝 결과와 정량데이터 분석 결과를 바탕으로 다양한 분야의 전문가들이 참여해 집단지성이 발휘 된다면 담당자의 직관에 의한 보정이라는 한계를 극복하고 예측력도 향상 될 수 있을 것이다.

셋째, 텍스트 마이닝을 활용해 산업의 트렌드를 파악하기 위해서는 시스템 사고를 기반으로 산업전반을 이해할 수 있는 카테고리별 연관관계도와 주제별연관도를 구축하고[9], 이러한 프레임워크를 바탕으로 관심 주제와 키워드를 도출해 지속적으로 정보를 수집하고 분석하는 체계를 가져간다면 더욱 트렌드 분석의 예측력은 높아질 수 있을 것이다[14].

본 논문의 한계와 향후 과제는 정보스크린 과정에서 단위정보들의 맥락(Context)과 히스토리(History)를 파악하기 위해 시스템 다이내믹스에 의한 철강산업 연관도분석을 보다 고도화는 노력이 필요하고, 지금까지 발견한 유의미한 트렌드와 전문가 의견을 반영해서 연관도와 카테고리별 주제와 키워드들을 지속적으로 보완해 나갈 필요가 있다.

트렌드 분석과정에서는 관련분야 전문가와 함께 토론을 거쳐서 텍스트 마이닝 툴에서 제공되는 데이터 패턴을 보다 의미 있게 해석해 내는 표준화된 프로세스와 방법론의 고민이 필요하다.

전문가 검증 과정에서는 전문가 의견이 위키피드식으로 온라인에서 트렌드 분석 리포트를 보완하는 체제도 고민해 볼 만하다.

마지막으로 정량데이터 분석 모델의 보완을 위해서는 핵심변수에 영향을 미치는 키워드를 도출, 텍스트 마이닝을 통해 변수의 추가 삭제 변경한다면 정량데이터 분석 모델의 설명력과 예측력을 높여 나가는 선순환이 이루어 질수 있을 것이다.

References

[1] Bae, J. H., Son, J. E., and Song, M., "Analysis of Twitter for 2012 South Korea Presidential Election by Text Mining Techniques," J Intell Inform syst 2013 September, Vol. 19, No. 3, pp. 141-156, 2013.

[2] Bae, S. J. and Park, C. K., "Analysis of text mining techniques Feasibility Study Technical Information," Korea Techno-

logy Innovation Society Conference 2003, pp. 75-88, 2003.

[3] Cho, G. H., Lim, S. Y., and Hur, S., "An Analysis of the Research Methodologies and techniques in the Industrial Engineering Using Text Mining," Journal of Korean institute of industrial engineers, Vol. 40, No. 1, pp. 52-59, 2014.

[4] Galit, S., Data mining for business intelligence, Wiley, 2011.

[5] He, Q., "Knowledge Discovery Through Co-Word Analysis," Library Trends, Vol. 48, No. 1, pp. 133-159, 1999.

[6] Jiawel, H., Data Mining Concepts and Techniques, Elsevier, 2011.

[7] Jung, Y. C., Bing Data, Communication books, 2013.

[8] Kim, B. H. and Hong, S. I., Principles of Economics, Chongmok Publisher, 2012.

[9] Kim, D. H., System Thinking, Sunhaksa, 2004.

[10] Kim, H. J. and Song, M., "A Study on the Research Trends in Domestic/International Information Science Articles by Co-word Analysis," Journal of the Korean society for information management, Vol. 31, No. 1, pp. 99-118, 2014.

[11] Kim, J. H. and Gong, M. G., "Structural Changes in the Global Steel Industry and Future Prospects," Journal of the Economic Policy and Industrial Research, Vol. 174, pp. 85-114, Fall 2012.

[12] Lee, S. J. and Kim, H. J., "Keyword Extraction from News Corpus using Modified TF-IDF," The Journal of Society for e-Business Studies, Vol. 14, No. 4, pp. 59-73, 2009.

[13] LG CNS Official Blog, Social analytics solutions using big data 'SMART SMA' intensive exploration, 2014, http://blog.lgcns.com/469.

[14] Ulf, P., Trends und Szenarien als Werkzeuge zur Strategieentwicklung, Publicis, 2008.

저 자 소 개

민기영 (E-mail : hoontae@daejin.ac.kr)
1992년 한양여자대학교 전자계산학과
2006년 연세대학교 공학대학원 컴퓨터공학과 (석사)
2012년~현재 연세대학교 일반대학원 기술정책박사 과정
2011년~현재 포스코경영연구소 재직
관심분야 빅데이터, 업무혁신, 프로세스 분석 및 설계

김훈태 (E-mail : yongguji@yonsei.ac.kr)
1989년 경북대학교 무역학과 (학사)
1991년 경북대학교 무역학과 (석사)
1992년~2000년 포스코경영연구소 연구원
2011년~현재 포스코경영연구소 수석연구원
2014년~현재 연세대학교 일반대학원 기술정책박사 과정
관심분야 빅데이터, 혁신전략

지용구 (E-mail : yongguji@yonsei.ac.kr)
1994년 서울대학교 산업공학과 (학사)
1996년 서울대학교 산업공학과 (석사)
2001년 Purdue University 산업공학과 (박사)
2002년~2005년 숭실대학교 정보산업공학과 조교수
2005년~현재 연세대학교 정보산업공학과 부교수
관심분야 HCI, Usability in IT Products