

소셜 빅데이터 이슈 탐지 및 예측분석 기술 동향

Technology Trends of Issue Detection and Predictive Analysis on Social Big Data

이충희 (C.H. Lee) 지식마이닝연구팀 선임연구원
허 정 (J. Hur) 지식마이닝연구팀 선임연구원
오효정 (H.J. Oh) 지식마이닝연구팀 선임연구원
김현진 (H.J. Kim) 지식마이닝연구팀 선임연구원
류법모 (P.M. Ryu) 지식마이닝연구팀 선임연구원
김현기 (H.K. Kim) 지식마이닝연구팀 팀장

2013
Electronics and
Telecommunications
Trends

빅데이터 처리 및
분석 기술 특집

- I. 머리말
- II. 소셜미디어 이슈 탐지
기술
- III. 예측분석 기술
- IV. 맺음말

최근 빅데이터 시대를 맞이하여 소셜미디어가 중요한 정보의 소통수단으로 급부상함에 따라 소셜웹 이슈 탐지 및 예측분석 기술이 큰 주목을 받고 있고, 기업·정부 등에서 정치/경제/사회문화적 이슈들에 대한 온라인 동향 분석 및 이슈 예측 기술의 수요가 급증하고 있다. 본고에서는 페이스북, 트위터 등의 소셜미디어에 대한 온라인 동향 분석 및 모니터링 기술 개발의 국내/국외 상용화 및 연구 현황을 소개한다. 또한, 사회적 동향을 분석해서 만들어진 예측모델에 기반해서 이슈의 향후 전개 과정에 대해 정량적으로 예측하는 기술 현황을 국내와 국외로 나누어 소개한다.

I. 머리말

단순한 소통 미디어였던 소셜미디어가 최근에는 트위터, 페이스북을 중심으로 활성화되면서 각 개인의 목소리가 모여서 커다란 목소리를 내는 발언대로 바뀌었고, 사회 전반적으로 화제의 중심에 서 있다. 그에 따라 소셜미디어 기반의 인터렉션(interaction) 데이터를 분석하여 각 분야에 활용하고자 하는 요구가 급증하고 있다.

소셜미디어 기반 데이터는 기존의 트랜잭션(transaction) 데이터보다 훨씬 방대하고 다양한 구조를 갖고 있지만, 각 사회 구성원들 사이에서 직접적인 상호 소통을 통해 생성·공유·교환되고 있기 때문에 현재 시점의 사회 동향 및 사회 구성원의 행동 패턴을 표현하고 있다고 인식되고 있다. 따라서 소셜미디어의 비구조적이고 복잡한 데이터를 이해하고 유용한 지식을 추출하여 현재의 주요 이슈를 탐지하고 모니터링하며, 미래를 예측하기 위한 연구가 활발히 진행되고 있다.

실제로 온라인 평판에 대한 공정한 평가방법론 및 평판 분석 기술의 필요성이 대두되고 있고, 소셜웹이 거대한 평판 시스템으로 진화함에 따라 온라인에서 좋은 평판을 획득하려는 개인, 기업 등의 공급자에 의한 온라인 평판의 쏠림현상/시장 왜곡이 빈번해지고, 그로 인해 사용자 불만이 높아지고 있다. 따라서 이를 극복할 수 있는 소셜웹 이슈 탐지-모니터링 기술이 필요해지고 있다. 또한 제품가격/판매량 예측, 영화 박스 오피스 예측, 선거 예측, 정보 확산도 예측, 거시경제 예측 등 다양한 분야에서 많은 연구 및 상용 서비스가 발표되고 있다.

본고에서는 이러한 사회적 현상에 발맞춰 국내와 국외에서 활발히 연구되고 상용화되고 있는 소셜미디어 기반 이슈 탐지 및 예측분석 기술 동향에 대해서 상세히 기술한다.

II. 소셜미디어 이슈 탐지 기술

1. 국내 기술 동향

국내에서는 2012년 총선을 기점으로 가장 큰 화두였던 대선에 이르기까지 소셜미디어상의 민심 동향을 보여 줄 수 있는 선거 관련 서비스 사례들이 두드러지고 있다. 주로 관련 정치인 트위터 실시간 업데이트, 정치 팟캐스트, 선거구 분석, 일반인들의 정치 관련 소셜미디어 동향 분석을 위주로 서비스를 구성하고 있다. 이러한 단기간 이슈에 국한된 서비스뿐 아니라 기존의 한국어 검색 서비스 업체들이 언어분석 기술을 접목하여 대용량 소셜미디어 분석으로 방향을 잡고 언어분석 기반 서비스를 출시하고 있다. 이러한 업체들이 제공하는 서비스는 크게 소셜미디어에서 검색어에 대해서 기간별로 출현 빈도를 모니터링하고, 해당 검색어와 관련 있는 연관어맵을 보여주고 있다. 또한 검색어와 관련된 문장에서 긍정과 부정의 감성을 분석하여 사용자에게 제공하고 있다. 국내에서 언어분석 기반으로 소셜미디어 분석 기능을 제공하고 있는 대표적인 서비스들을 소개하면 다음과 같다[1].

가. 소셜메트릭스(다음소프트)

소셜메트릭스(<http://www.socialmetrics.co.kr>)는 국내 텍스트마이닝 전문 업체인 다음소프트가 자연어처리 기술과 텍스트마이닝 기술을 바탕으로 블로그, 트위터 문서를 분석한 모니터링 정보를 제공하는 서비스이다. 소셜메트릭스는 사용자가 입력한 키워드에 대해 소셜미



(그림 1) 소셜 메트릭스 소셜검색 서비스 화면

디어에서 노출된 빈도 추이와 관련 연관어맵을 제공하고, 각 키워드에 대한 긍부정 감성정보 및 날짜별로 가장 많이 확산된 트윗 메시지를 제공하고 있다(그림 1) 참조).

나. 트루스토리(솔트룩스)

기업용 검색 솔루션 개발업체인 솔트룩스는 정치인 혹은 정치적 관심을 받는 인물들을 대상으로 소셜미디어상에서의 추이를 제공하는 트루스토리(<http://politician.truestory.co.kr/main>) 서비스를 제공한다. 소셜미디어상에서 특정 분야의 이슈에 초점을 맞추는 차별화를 갖고 있는 이 서비스는 이슈화되고 있는 특정 인물들을 비교해서 보여주는 등 정치 분야에 특화된 기능을 제공하고 있다(그림 2) 참조).

다. 펄스K(코난테크놀로지)

검색전문업체인 코난테크놀로지는 트위터, 페이스



(그림 2) 트루스토리 '정치인' 관련 이슈 서비스 화면



(그림 3) 펄스K 모니터링 서비스 화면

북, 미투데이 및 블로그 등의 소셜미디어를 모니터링하고 분석할 수 있는 펄스K(<http://www.pulsek.com>) 서비스를 제공하고 있다. 펄스K에서는 최근 14일간의 소셜 인지도와 호감도(긍부정)를 기반으로 평판을 수치화하고, 기간별로 추이를 모니터링하는 기능을 제공한다(그림 3) 참조).

라. 씨날(그루터)

오픈소스인 Hadoop을 활용한 클라우드 기반 플랫폼 제공업체인 그루터는 이러한 Hadoop 기반 플랫폼을 기반으로 기업 및 브랜드 관련 소셜 베즈 모니터링 서비스인 씨날(<http://www.seenal.com>)을 출시하였다. 기존 소셜미디어 분석 서비스가 분석 성능과 서비스 차별화를 목표로 가렸다고 본다면, 씨날은 클라우드 기반 플랫폼을 활용한 효율적인 소셜미디어 운영과 소셜미디어 확산 경로 파악 등으로 기존 경쟁업체들과의 차별성을 내세우고 있다(그림 4) 참조).

마. 소셜위즈덤(한국전자통신연구원)

한국전자통신연구원에서 개발 중인 소셜미디어 이슈 탐지 및 모니터링 플랫폼(이하 소셜위즈덤)은 소셜미디어(뉴스, 블로그 트위터) 문서를 수집/저장하고, 차별화된 심층 언어분석을 기반으로 추출한 정보를 이용하여 이슈를 탐지하고 모니터링하는 기능을 제공한다. 소셜



(그림 4) 씨날의 리트윗 지도 서비스 화면

위즈덤은 문서 수집/저장/필터링 모듈, 언어분석 모듈, 정보 추출 모듈, 이슈 탐지 모듈, 이슈 모니터링 모듈과 관련 정보의 연계를 정의한 이슈 지식베이스로 구성된다.

기존 소셜미디어 분석 서비스들에서는 대부분 빈도를 기반으로 이슈어를 선정하기 때문에 항상 많이 나타나는 단어나 계절적, 이벤트성으로 정기적으로 출현 빈도가 높아지는 단어에 대해서는 진정한 이슈성을 계산하기 힘든 반면, 소셜위즈덤 서비스에서는 이러한 취약점을 보완하는 이슈 모델링을 제시하고, 이를 바탕으로 빈도와 별도로 이슈성을 반영한 이슈 탐지 방법론을 적용했다.

특히 소셜미디어의 감정을 표현한 문장들에 감성분석 기술을 접목하여 전반적인 호감도나 여론 등을 추론하는 서비스를 제공하고 있는데, 기존 서비스들이 긍부정 두 가지 감성만을 분석하는 반면, 소셜위즈덤에서는 긍정과 부정의 양극성을 좀 더 세분화한 감성분석 결과를 제공한다. 이러한 세부분류 감성분석 기술은 긍정과 부정의 전체적인 의견분석뿐 아니라, 긍정의견 중에서도 '감동', '만족' 등의 세부감성별로 의견을 분류해 볼



(그림 5) 소셜위즈덤 감성분석 결과

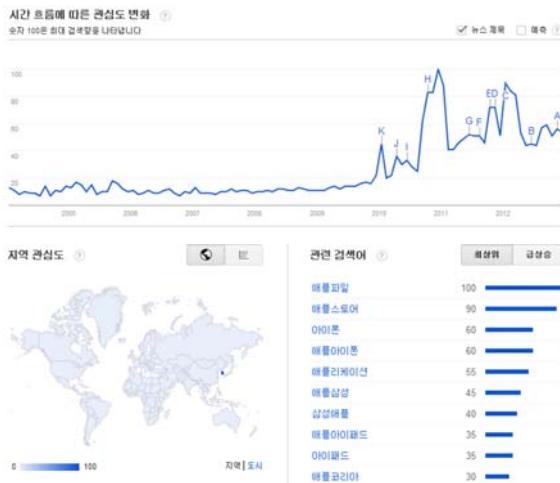
수 있고, 부정의 의견 중에서도 '싫어함'인지 '화남'인지에 따라서 이슈가 발생했을 때 대처하는 방향이 달라질 수 있으므로 보다 정확한 모니터링을 할 수 있는 장점이 있다. (그림 5)는 '아이폰'에 대한 긍부정 감성분석 결과이다.

2. 해외 기술 동향

해외에서는 블로그가 활성화된 2000년 이후부터 소셜미디어에서 사용자의 니즈(needs)를 파악하기 위한 비즈니스 연구가 진행되어 왔다. 특히 비즈니스 인텔리전스(Business Intelligence: BI) 분야에서는 특정 브랜드의 모니터링, 사용자의 피드백 감성분석 및 네트워크 분석 기반 구전 마케팅 및 의사결정을 지원하는 제품들이 개발되었다. 또한 각국에서 공익적인 측면에서 다양한 웹 및 소셜웹으로부터 질병과 재해 관련 정보를 탐지하여, 세계적 동향을 분석하고, 모니터링하는 서비스를 제공하는 연구도 활발히 진행 중이다. 최근에는 소셜웹의 중요성이 부각되면서 IBM, Microsoft, SAS 등 IT 대기업들이 비즈니스 인텔리전스(BI), 고객관계관리(Customer Relationship Management: CRM) 분야에 소셜웹 분석을 활용하는 제품들을 출시하고 있다. 대표적인 연구 사례는 다음과 같다.

가. Google Trends(미국, 구글)

구글은 전 세계에서 매일 수집되는 대량의 웹 사용자의 검색 로그를 기반으로 한 동향 분석 서비스(<http://www.google.com/trends>)를 제공하고 있다. 뿐만 아니라 각 개인 관심사와 구글 검색 로그 기반 전 세계인 관심사와의 비교가 가능하며, 사용자의 관심 주제의 시간상의 변화 관찰, 지역별 사용자 관심 추이 관찰이 가능하다(그림 6) 참조). 나아가 구글에 입력된 사용자 검색어가 특정 산업의 현 경제활동 상황과 상관관계가 있다고 가정, 다양한 예측모델을 제시하는 서비스도 제공하고 있다.



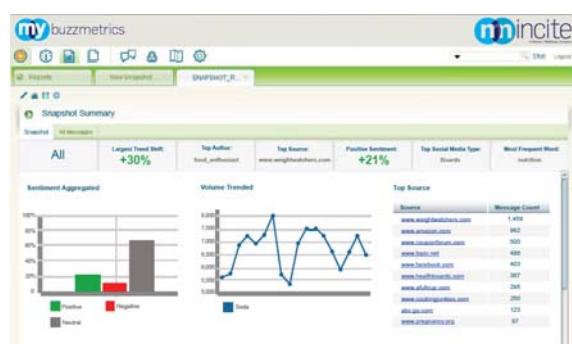
(그림 6) Google Trends 서비스 화면

나. BuzzMetrics(미국, 닐슨)

세계 선도적인 브랜드 모니터링 서비스를 제공하는 글로벌 시장조사 업체인 닐슨은 버즈메트릭스(BuzzMetrics, <http://nmincrite.com/>)라는 제품으로 기업, 고객에 대한 위기 탐지, 대응, 후속 조치에 대해 통찰력 있는 정보를 제공하며, 브랜드의 이미지나 고객의 반응을 BRM(Brand Association Map)으로 시각화하여 (그림 7)과 같은 형태로 제공한다. 또한 사용자의 구전(buzz)에 대한 감성분석을 이용하여 호불호 등을 제공함으로써 고객대응에 필요한 마케팅 전략을 수립하는 데 도움을 준다.

다. MediSys(유럽연합, Joint Research Center)

메디시스(<http://emm.newsbrief.eu/overview.html#explorer>)



(그림 7) BuzzMetrics 서비스 화면

medisys)는 유럽연합(EU)에서 진행 중인 의학 관련 이슈 탐지 프로젝트로, 250개의 의학전문 사이트와 400~1,600개의 일반 뉴스 포털, 20개의 유럽 지역 상업 뉴스 서비스에서 제공하고 있는 웹 미디어를 대상으로 의학 및 공공 건강 관련 주제를 지역 및 카테고리별로 분석하여 매 20분마다 자동으로 생성되는 조기 경고 서비스를 제공한다.

라. EMM NewsExplorer(유럽연합, FP-7)

유럽연합에서 지원하는 표준화 프로젝트(FP-7) 중 하나인 EMM NewsExplorer (<http://emm.newsbrief.eu/overview.html#explorer>)는 19개의 언어를 통해 일단위로 뉴스를 요약해주며, 사람/기관/나라에 관한 정보를 수집, 같은 사건이나 같은 주제의 뉴스를 군집화해서 연결함으로써 특정 사건의 군집들이 시간상 변화하는 트렌드를 탐지하는 기술을 연구하고 있다.

마. Radian6(미국, Radian)

미국에서 가장 널리 보급된 리스닝 플랫폼을 제공하는 기업인 래디안은 전통적인 기업 리스닝 솔루션을 위한 다양한 대시보드 기능, 영향력자 판별 기능을 제공하는 래디안6(<http://www.radian6.com>)를 선보이고 있다. 특히 최근에는 트위터, 페이스북 등 소셜웹 참여 기능과 소셜웹에 나타난 감성을 분석하는 기능을 함께 제공함으로써 온/오프라인 리스닝 솔루션을 표방하고 있다.

III. 예측분석 기술

1. 국내 기술 동향

예측분석과 관련해서 국내는 주로 비즈니스 인텔리전스(BI) 솔루션들을 중심으로 정형적 데이터 또는 수치 데이터를 이용한 데이터 분석이 주를 이루며, 일부 연구 기관에서 전문가의 예측을 지원하는 도구개발을 시도하

고 있지만 국외 기술에 비해 시작 단계이고 수작업 의존도가 높다. 또한 국내에서는 주로 수치 데이터를 이용한 예측분석이 주를 이루며, 비정형적 데이터에서 정보를 추출한 예측연구는 시작단계이다. 대표적인 연구사례들은 다음과 같다.

가. 계량정보분석

정보자원의 기하급수적 증가에 따라 필요한 정보만을 선별하여 분석하고 전략적으로 활용하는 것이 점점 더 중요해지고 있으며, 그에 따라 계량정보학이 급속히 연구되고 있다. 계량정보학은 과학문헌을 대상으로 어떤 특성을 도출하기 위한 수단으로써 과학문헌을 정량적으로 표현하는 연구 분야이다.

국내 KISTI에서 개발한 KnowledgeMatrix는 논문 및 특허의 서지정보를 분석하여 지식을 발견하기 위한 목적으로 설계된 계량정보분석 시스템이며, 대규모 정보를 구조화/계량화하고 기술의 트렌드, 급부상 연구 분야, 연구자 네트워크 등을 분석 및 시각화해서 보여줄 수 있다[2]. KnowledgeMatrix의 주요 특징은 다음과 같다.

- 논문, 특허 분석을 통한 미래유망기술 후보 및 백데이터 제시
- 피인용율이 높은 논문 및 특허를 분석하여 과학 기술 패러다임 추적
- 특정 기술에 대해 세부기술별로 유망, 성장, 쇠퇴, 도래할 연구 분류
- 최종 단계에서는 전문가가 개입하여 유망기술 선정
- 명사구 추출 수준의 간단한 텍스트 분석만 사용

나. 과채류 가격 예측

농산물 가격 예측은 주로 채소류, 과일류를 중심으로 수급전망을 통한 연도별 가격 예측과 시계열 분석을 통한 월별, 연도별 가격 예측 방법이 이용되어 왔다.

한국농촌경제연구원에서는 과채류 3개 품목 6개 품종

을 대상으로 VAR 모형, ARMA 모형, GARCH 모형을 이용하여 단기적 가격 예측을 행하였고, 추정된 예측치와 실측치를 비교하여 각 품목 품종에 적합한 가격 예측 모형을 찾는 연구를 수행하였다[3]. 연구결과에 의하면 백다다기오이, 취청오이, 일반토마토, 방울토마토에 대한 가격 예측에서는 VAR 모형보다 ARMA 모형이 우수한 결과를 보였고, 애호박, 쥬키니 호박은 VAR 모형에서 개선된 예측치를 보였다.

다. KTX 수요 예측

효율적인 철도 운영을 위해서는 각종 의사결정의 기초자료인 철도수요를 예측하는 것이 필요하다. 철도수요의 예측방법은 크게 과거 수송실적에 기반한 패턴인식 방법과 열차운행 횟수, 운임 등의 정책 변화를 고려 할 수 있는 회귀모형, 직접수요모형, 전통적 4단계 모형으로 구분할 수 있다.

한국철도공사에서는 패턴인식 방법 중 시계열 모형에 기반해서 KTX 수요를 예측하기 위한 방법으로 개입 ARIMA 모형을 제안하였다[4]. 이 모형은 다음과 같은 자료에 기반해서 예측모형을 설계하였다.

- KTX 통행량을 주중과 주말로 구분하여 예측
- KTX 통행량의 월별 계정변동 반영
- 경부고속철도 2단계 개통 등과 같은 개입 반영

모형식별과 모형추정은 충분한 관측 개수가 필요하므로 각각 79개월, 84개월을 분석 기간으로 설정하였다. KTX 통행량은 2004년 4월 개통 이후 2006년까지 증가하다, 2010년 11월 경부고속철도 2단계 개통으로 인해 한 단계 증가하는 추세를 나타내고 있으며, 개입 ARIMA 모형은 이와 같은 개입이 존재할 경우 효과적인 예측이 가능하였다.

2. 해외 기술 동향

사회적 이슈에 대한 분석 및 예측 기술은 사회/문화적 다양한 위험을 사전에 감지하고 대처할 수 있어서 사회

/문화적 위험 대처를 위한 간접비용을 줄일 수 있으며, 삶의 질을 향상시킬 수 있는 중요한 기술로 인지되고 있다. 이로 인해 국외에서는 다양한 자원을 활용한 사회적 동향 분석 및 위험 감지에 대한 연구를 수행하고 있다. 특히, 선거, 경제문제(실업, 주택), 질병, 자연재해 등과 같은 사회/경제적 위험 요소에 대한 예측과 소셜미디어

의 여론 동향 분석이 비교적 용이한 선거 결과 분석 및 예측에 많은 연구가 진행되고 있다. 가트너에서 2010년에 발표한 분석 응용 기술에 대한 hyper cycle에 'Predictive Campaign Analytics' 기술과 'Real-time Decision' 기술이 차세대 주요한 기술임을 전망하였다.

〈표 1〉은 국외 소셜미디어 예측분석 기술에 대한 주

〈표 1〉 국외 소셜미디어 예측분석 주요 연구내용

연구제목	기관/년도	정보 소스	연구내용
Predictive Search Engine (Temporal & Predictive Analytics)	Recorded Future (미국), 2009	웹 사이트 (뉴스, 블로그, 트위터, 공공 데이터, 무역 출판물, 정치 경제적 연설 등)	<ul style="list-style-type: none"> 실시간 스트리밍 데이터를 분석하여 과거/현재에 일어날 이슈에 관한 힌트를 발견하는 둘 제품 발표, 자연재해, 회사 합병 등의 사건과 예상 시기, 관련된 객체(사람, 회사, 나라 등) 탐지. 함께 추출된 시간 정보가 미래인 경우, 미래의 예측 근거로 제시(실제 예측 기능을 제공하지는 않음) 객체에 관한 뉴스상에서의 감성 분석 Google과 CIA로부터 펀드 지원 받음. https://www.recordedfuture.com/
Predicting the Present with Google Trends	Google(미국), 2009	Google Trends/Insights	<ul style="list-style-type: none"> Google 검색어가 특정 산업의 현 경제활동 상황과 상관관계가 있다고 가정함. Retail sales, automotive sales, home sales, travel 분야에서 예측모델 제시
Predicting Initial Claims for Unemployment	Google(미국), 2009	Google Trends/Insight	<ul style="list-style-type: none"> 미국에서 실업 수당 청구 건수 예측
Forecasting US Home Foreclosures	San Jose State Univ.(미국), 2009	인터넷 키워드 서치의 인덱스	<ul style="list-style-type: none"> 미국 주택 시장에서 주택 차압 수의 단기적 예상 모델
Google Flu Epidemics Reporting	Google(미국), 2009	Google 검색 로그	<ul style="list-style-type: none"> 감기 관련 검색어의 빈도를 계산하여 전 세계 여러 국가, 지역에서 독감 유행 수준을 예측 http://www.google.org/flutrends/
Predicting the Future with Social Media	HP(미국), 2010	Twitter	<ul style="list-style-type: none"> 특정 영화에 관해 오가는 tweet 률에 근거한 영화 흥행율을 예상하는 모델 제안 마켓 기반의 예상보다 적중률이 높음.
Predicting Elections with Twitter	Tech. Univ. Munchen (독일), 2010	Twitter	<ul style="list-style-type: none"> 2009 독일 연방 선거 대상 당이나 입후보자에 관한 tweet의 수만으로 유권자의 선호도 판별 가능, 전통적 방식의 여론 조사와 유사한 결과
Prediction Markets	Univ. of Columbia, Univ. of Pennsylvania(미국), 2008	News Articles	<ul style="list-style-type: none"> 뉴스 미디어에 다루어지는 (객관적) 사건들이 공공의 의견에 어떤 영향을 미치는지 연구 미국 대통령 선거 대상
Toretter	도쿄대(일본), 2009~2010	소셜웹 (Twitter)	<ul style="list-style-type: none"> 실시간 지진 리포팅 시스템 지진의 확산 속도보다 빠른 경보 가능 http://toretter.com
Predictive Opinion: Crystal	Univ. of Southern California(미국), 2007	웹사이트	<ul style="list-style-type: none"> 선거 결과 예측 웹 사이트에 올라온 사용자의 의견을 분석, 결과 예측

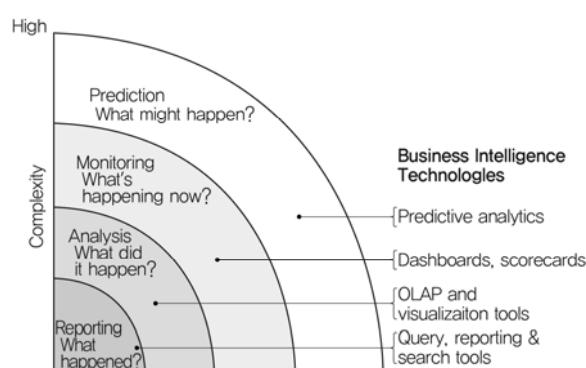
요한 연구 및 내용을 정리한 것이다. 앞서 언급한 바와 같이 대부분의 연구는 사회/경제적 위협요인에 대한 예측과 선거를 대상으로 하고 있다.

가. 경영정보학 분야의 기술 동향

최근 실시간으로 생성되는 엄청난 양의 소셜미디어 정보와 기존에 구축되어 있던 다양한 형태의 데이터 정보를 대상으로 한 정보분석 및 통합에 대한 요구가 급증하면서 빅데이터 분석 기술에 대한 관심이 폭발적으로 증가하고 있다. 특히 경영정보학(BI) 분야에서는 기업 브랜드 및 상품 홍보와 기업 전략 기획 등과 같은 경영 전반의 의사결정을 위해서 정보 리포팅, 분석, 모니터링, 예측 등의 기술이 활용되고 있다. (그림 8)은 경영정보학 분야의 기술 스펙트럼으로써, 기업 입장에서 볼 때, 가치가 가장 높은 기술이 예측(prediction)이지만 기술적 난이도도 가장 높은 것을 알 수 있다. 결국 빅데이터 분석 기술의 핵심은 대용량의 데이터에 대한 정보분석을 기반으로 경영 전반에 대한 중요 이벤트의 분석 및 예측이 될 것이고, 이를 기반으로 최적의 의사결정에 활용하는 것이 경영정보학의 목적이다.

가트너에서 발표한 “Forecast: Global Business Intelligence Software Market, 2009~2014”에 의하면 예

〈자료〉: TDWI Checklist Report: Big Data Analytics, 2010.



(그림 8) The Spectrum of BI Technologies

측분석(predictive analytics)과 데이터마이닝(data mining) 시장의 성장이 점차적으로 증가하는 것을 알 수 있다. 이처럼 빅데이터 분석과 소셜미디어 이슈 예측분석 기술은 시장 잠재력이 풍부하다.

나. 사회 분야의 기술 동향

2009년 유엔에서 인간의 복지와 관련된 다양한 변화와 영향을 정확하게 이해하고 위기에 대응하기 위해 데이터 분석을 토대로 한 전략을 수립하기 위해 UN Global Pulse라는 기관을 창설하였다[1]. UN Agency, 정부, 학계, 민간 등의 전문가로 구성되었으며, 〈표 2〉와 같은 연구를 수행하였다. 각 연구의 세부적인 내용은 다음에서 자세히 설명한다.

1) 트위터 분석을 통한 위기에 대한 스트레스 인식 조사

인도네시아어/자바어/영어로 작성된 공개 트윗을 대상으로 “글로벌 경제위기에 사람들이 대처하는 방법을 알 수 있는지표가 소셜미디어에 존재하는지”와 “위기에 대한 사람들의 스트레스 인식을 조사하기 위해 트위터의 활용 가능성과 방법론이 있는지”에 대해서 하버드 대학교 정량 사회 과학 연구소(Institute for Quantitative Social Science)의 개리 킹(Gary King) 교수가 분석하였다. 분석 결과, 트위터 데이터는 위기에 대한 사람들의 감정을 이해하는 데는 유용하지만, 장기적 이슈에

〈표 2〉 UN Global Pulse의 수행 프로젝트

수행 프로젝트	설명
Future of Real-Time	실시간 데이터 스트림 분석을 통해 사회 및 개인 안전 관련 주요 이머징 트렌드 도출
Unemployment Through the Lens of Social Media	소셜 데이터 분석을 통해 취업 관련 주제, 분위기 등을 조사하고 공식적인 실업 통계와 비교
Monitoring Food Security Issues Through News Media	8년간의 뉴스 정보를 텍스트 분석, 시맨틱 클러스터링, 네트워크 이론 등을 적용하여 식품 안전성 관련 이머징 트렌드 도출

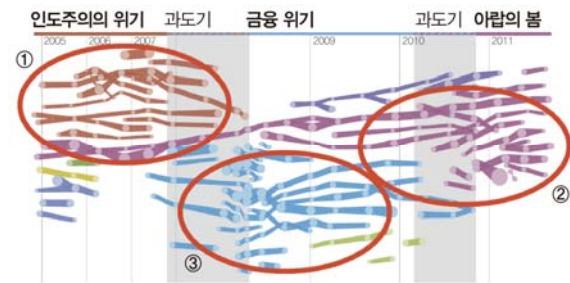
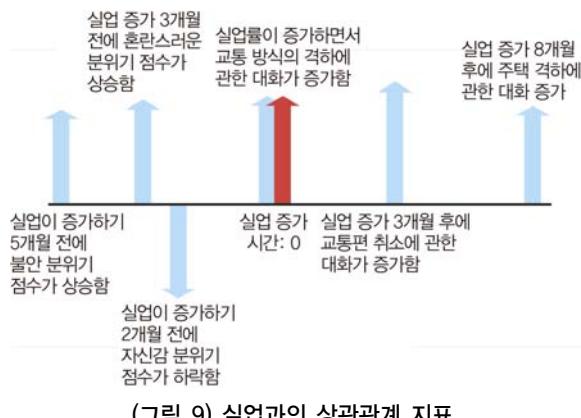
대한 판단에는 데이터가 충분하지 않다는 것을 알 수 있었다. 또한 분석의 효과를 극대화하기 위해서는 트위터 고유 문화, 시공간적 차이, 분석 대상의 구체성과 포괄성의 균형이 중요함을 알 수 있었다.

2) 소셜미디어와 온라인 대화를 이용한 실업 통계 심층 분석

소셜미디어에 나타나는 사람들의 감정 및 정보와 실제 실업률 변화와의 연관관계를 파악하기 위해서, 미국과 아일랜드를 대상으로 문서를 수집하여 분석하였다. 분석 단계는 6가지의 감정 카테고리에 따른 감정분석과 선행지표, 후행지표를 찾기 위한 교차 상관관계 분석으로 나뉘어 진행되었다. (그림 9)는 교차 상관관계 분석에 따른 실업률 관련 상관 지표이다. 이처럼 소셜 분석은 실업에 대한 선/후행 지표를 기반으로 향후 실업률의 상승 및 하락에 대한 예측과 실업률 증가에 따른 사회적 영향을 예측할 수 있다. 또한 실업증가에 따른 사회적인 정책대안을 마련함으로써 효율적인 실업정책을 수립할 수 있다.

3) 미디어에 나타난 이슈 흐름 분석을 통한 세계 식량 안전 모니터링

디지털 데이터 분석을 통해 21세기 식량 안전 관련 이슈의 변화와 세계의 지역별 특성을 파악하여 식량 정책 분석의 가치를 중대시킬 목적으로 연구가 진행되었



(그림 10) 하이퍼 스트림의 변화 분석

다. 약 2,000개의 뉴스 기사를 대상으로 어휘를 추출하고, 어휘들 간의 관계성을 기반으로 클러스터를 생성한다. 생성된 클러스터들은 관계성에 따라 스트림을 통해 연결하며 유사한 주제를 가진 스트림은 같은 색깔로 그룹핑하여 하이퍼 스트림으로 정의하였다. (그림 10)에서 빨강 스트림(①)은 ‘특정 지역의 기아는 비효율적인 식량 생산과 분배에 기인한다’는 지역 단위의 현상을 표현하고, 분홍 스트림(②)은 ‘식량불안은 정치적 혼란과 자연재해로 발생한다’는 인도주의적 관점의 표현이다. 연파랑 스트림(③)은 식량 위기는 글로벌 경제 정세에 기인하는 것을 표현한다.

IV. 맺음말

소셜미디어가 현재 다양한 분야에 영향력을 발휘하고 있고, 소셜미디어의 효과가 다양한 연구 발표로 입증되고 있다. 그러므로 소셜미디어로부터 쏟아져 나오는 방대하고 복잡한 소통 데이터를 적극적으로 활용하려는 시도가 정부, 기업, 연구소, 대학 등을 중심으로 다각적으로 이루어지고 있다.

소셜미디어 분석을 통한 이슈 탐지의 중요성이 매우 커지고 현재 다양한 연구들이 이루어지고 있지만, 아직은 선거나 기업 제품 등의 일부 분야에 대해서만 이루어지고 있고, 정보의 쏠림현상/정보왜곡 등의 문제가 여전히 남아있다. 또한 예측분석의 중요성과 필요성은 정부/기업/대학 등에서 공감하고 있지만 아직은 시작 단

계로 대용량 데이터의 실시간 분석을 통한 예측 및 장기적인 예측 방법에 대한 연구가 예측분석의 나아갈 방향이라고 전망한다.

용어해설

소셜미디어 개방, 참여, 공유의 가치로 요약되는 웹 2.0 시대의 도래에 따라 소셜 네트워크의 기반 위에서 개인의 생각이나 의견, 경험, 정보 등을 서로 공유하고 타인과의 관계를 생성 또는 확장시킬 수 있는 개방화된 온라인 플랫폼을 의미함.

소셜웹 뉴스, 블로그, SNS(트위터, 미투데이 등), 게시판, 카페 등을 모두 포괄해서 나타내는 용어

소셜웹 이슈 탐지/모니터링 소셜미디어로부터 이슈의 징후를 탐지하고 전개 과정을 모니터링하는 기술로, 모니터링 대상 이슈의 종류는 지속적으로 추적이 필요한 지속적 이슈, 일정 주기로 추적이 필요한 반복적 이슈, 단발적으로 나타나는 일회적 이슈로 구분함.

이슈 예측분석 과거의 데이터를 계량적 방법(예: 회귀분석, 시계열 분석 등), 질적접근 방법(예: 델파이 방법) 등을 통해 예측모델을 추정하고 현재의 데이터와 예측모델을 이용하여 이슈의 향후 전개 과정에 대한 예측모형을 분석하는 기술

약어 정리

SNS	Social Network Service
BI	Business Intelligence
CRM	Customer Relationship Management
BRM	Brand Association Map

참고문헌

- [1] 류법모 외, “심층 언어분석 기반 소셜미디어 이슈 탐지 및 모니터링 기술,” *정보과학회지*, vol. 30, no. 6, 2012, 5, pp. 47-58.
- [2] 이방래 외, “계량정보분석시스템 KnowledgeMatrix 개발,” *한국콘텐츠학회 종합학술대회 논문집*, vol. 5, no. 1, 2007, pp. 167-171.
- [3] 최병우, 최의창, “시계열 분석방법을 이용한 과채류 월별가격 예측,” *농촌경제*, vol. 30, no. 1, 2007, pp. 129-148.
- [4] 김관현, 김한수, “개입 ARIMA 모형을 이용한 KTX 수요 예측,” *한국철도학회 논문집*, vol. 14, no. 5, 2011, pp. 470-476.