

소셜미디어에 나타난 연속성 이슈 이벤트 분석

오효정[†] · 김현기^{††} · 윤보현^{†††}

요 약

본 논문에서는 다양한 소셜미디어 채널을 통해 수집된 시시각각 발발하는 사건/사고들 중, 연관된 사건이 시간과 장소적으로 어떠한 연속성이 나타나는지를 분석하고자 한다. 특히 사회적으로 파급력이 큰 사건/사고를 ‘이슈 이벤트’라고 정의, 도메인별로 발생하는 연속성 이벤트 유형과 특성을 규명한다. 또한, 소셜미디어 상에서 발생한 연속성 이벤트를 자동으로 탐지, 이를 추적하는 이슈 이벤트 연속성 분석 시스템에 대해 기술하며, 이를 통해 추출된 특정 도메인에 대한 이슈 이벤트들의 연속성을 시간 및 장소 축으로 도식화하여 분석하고, 나아가 이슈 이벤트의 전파 양상을 통해 소셜미디어 간의 상관관계를 파악한다.

주제어 : 연속성 이벤트, 이벤트 추출, 시공간 분석, 소셜미디어

Continuous Issue Event Analysis in Social Media

Hyo-Jung Oh[†] · Hyunki Kim^{††} · Bo-Hyun Yun^{†††}

ABSTRACT

This paper reveals continuity of related events which are occurred and changing from moment to moment accident/events collected from various social media channels. Among them, we especially define the events which have big social influence as “issue event” and investigate the type and characteristics of continuous issue event for each domain. We also introduce a automatic issue detection system in social media text. Based on the extracted issue event results in a particular domain, we analyse the continuity of those events by illustrating in time and place-axis. Furthermore, we identify the relationship between social media in terms of issue events propagation.

Keywords : Continuous event, Event extraction, Geo-temporal analysis, Social media

[†] 정 회 원: 한국전자통신연구원 지식마케팅연구소 선임연구원 (주저자)

^{††} 정 회 원: 한국전자통신연구원 지식마케팅연구소 팀장

^{†††} 정 회 원: 목원대학교 컴퓨터교육과 (교신저자)

논문접수: 2013년 7월 29일, 심사완료: 2013년 9월 27일, 게재확정: 2014년 1월 13일

* 본 논문은 2013년 미래창조과학부 및 한국산업기술평가관리원의 산업융합원천기술개발사업(정보통신)의 일환으로 수행하였음. [과제번호: 10044577, 휴먼 지식증강 서비스를 위한 지능진화형 WiseQA 플랫폼 기술 개발]

* 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임. (No. 한국연구재단에서 부여한 과제번호 : 2010-0022201)

1. 서론

웹 2.0 시대의 개막과 함께 과거에 뉴스를 전달하는 매체였던 인쇄물이 웹의 형태로 진화하고 있으며, 최근에는 오직 웹을 이용하여 뉴스를 제공하는 인터넷 신문 사이트도 다수이다. 나아가 개인이 각자의 논평을 덧붙여 1인 신문사가 되는 소셜 뉴스 서비스도 등장하는 추세이다.

뿐만 아니라 트위터나 페이스북과 같은 소셜미디어 서비스의 폭발적인 인기에 힘입어, 개인이 직접 목격하고 경험한 사건/사고 내용의 실시간 전파가 가능해짐에 따라 소셜미디어가 기존 전통적인 언론사보다 더 빠르고 파급력 있는 뉴스보도의 역할도 겸하고 있는 실정이다[1].

본 논문에서는 다양한 채널을 통해 수집된 시시각각 발발하는 사건/사고들 중, 연관된 사건이 시간과 장소적으로 어떠한 연속성이 나타나는지를 분석하고자 한다. 특히 사회적으로 파급력이 큰 사건/사고를 ‘이슈 이벤트’라고 명명하고자 한다.

일반적인 의미의 ‘이벤트’는 문제가 되거나 관심을 끌만한 일을 말하는 반면, 디지털 정보처리를 위한 정보추출(Information Extraction) 관점에서의 ‘이벤트’는 사용자 혹은 주어진 주제에 해당하는 ‘사건’에 해당하는 것으로, 정보추출 대상(target)을 의미한다. 특히 본 논문에서는 이벤트의 특성에 따라 단발성 이벤트와 연속성 이벤트로 분류하였다. ‘단발성’ 이벤트란 자동차 사고나 도난 사고 등과 같이 특정 이벤트가 발생을 하여도 타 지역 혹은 다른 시간대에 발생한 유사 사건과 연관성이 없는 이벤트를 의미한다. 반면, 연속성 이벤트는 감염성 질병이나 태풍과 같이 최초 이벤트가 발생한 뒤 시간 흐름에 따라 인접 지역으로 동일한 이벤트가 확산되는 이벤트를 말한다. 단발성 이벤트에 비하여 연속성 이벤트가 사회적으로 미치는 영향이 크기 때문에 온라인 콘텐츠 상에서 발생한 연속성 이벤트를 자동으로 탐지, 이를 추적할 수 있다면 이벤트의 최초 발생 이후의 이벤트 발생 경로와 확산 범위의 분석이 가능해지며, 이를 통해 신속하고 효과적인 대응 방안을 수립하는데 도움이 된다.

본 논문의 주안점은 도메인별로 발생하는 연속성 이벤트 유형을 정의하고, 해당 이벤트들의 연속성을 시간 및 장소 축으로 분석하는데 있다. 또한 미디어 유형별로 발생된 이벤트의 전파 양상을 파악하기로 한다. 논문의 구성은 다음과 같다. 2장에서는 연속성 이벤트와 관련한 기존 연구에 대하여 살펴보고, 3장에서는 소셜미디어에 나타난 이벤트들을 조사한 결과를 바탕으로 연속성 이벤트의 특성을 규명한다. 4장에서는 이슈 이벤트 연속성 분석을 위한 시스템을 소개하고, 5장에서는 실제 추출된 특정 도메인에 대한 이슈 이벤트 연속성을 시간 및 장소 축으로 도식화하여 분석하고, 나아가 소셜미디어간의 상관관계를 파악한다. 마지막으로 6장에서는 결론과 향후 연구에 대해 언급한다.

2. 관련 연구

그날 처음 발생한 사건, 또는 같은 사건을 다루는 기사들을 탐색하고, 예전에 발생한 사건과 관련된 사건인지를 추적해 나가는 기술을 사건 탐색 및 추적 (Topic Detection and Tracking, 이하 TDT) 이라고 한다[2][3]. TDT 이벤트[3]는 이전 이벤트와 관련하여 어떤 사건이 특정 시간 혹은 특정 장소에서 발생하는 이벤트를 말하며 예를 들어 선거 운동, 재난 복구 과정 등이 있다. 초창기의 TDT 기술[5]은 주제의 형태에 따라 문서를 군집/분류하는 수준에 머물렀지만, [6]에서는 개체명과 명사, 명사 구의 공기 빈도를 활용하여 이벤트를 구축하여 이벤트를 탐지하였다. 또한 [7]에서는 DFT(Discrete Fourier Transformation)와 각 카테고리 별로 분류된 자질을 사용하여 이벤트를 추적하였다. TDT 기술은 스트림으로 입력되는 데이터에서 자동으로 새로운 이벤트를 탐지하거나 입력된 쿼리(query)나 키워드(keyword)에 해당하는 주제를 탐지하는 것이 대부분이라 연속성 이벤트의 확산 유형 및 확산 범위를 분석하는 본 연구와는 차이가 있다.

한편, 모바일 단말의 보편화로 인해 소셜미디어가 폭발적 증가하는 추세에 힘입어 최근에는 사용자의 위치 정보를 제공되는 트위터 데이터를 활용하는 연구가 진행되고 있다. 약 3년치 트위터

데이터를 분석하고 감정적인 단어를 추출하여 지역 간의 분위기(mood) 변화를 분석하는 연구[8]가 진행되었고, 트위터의 실시간으로 업데이트 되는 특성과 다양한 주제를 허용하는 포괄적인 특성을 활용하여 지진과 같은 자연 재해 현상을 때때로 뉴스 미디어에서 제공하는 정보보다 더 빠르게 활용될 수 있다는 연구[9][10]가 제기되었다. 본 논문에서는 소셜미디어 중 뉴스와 트위터와의 이벤트 전과 경로, 연관관계 등을 분석하기로 한다.

3. 연속성 이슈 이벤트 특성 조사

소셜미디어에 나타난 연속성 이슈 이벤트의 특성을 파악하기 위해, 먼저 특정 기간 내의 뉴스 미디어를 분석하였다. 2010년 11월부터 2013년 2월 사이의 웹 뉴스를 수집하여, 시간 및 장소의 연속성이 유지되는 이슈 이벤트 선정에 위한 기초 조사를 실시하였다. 그 결과, 주로 ‘보건/의료’ 관련 뉴스기사에서 ‘질병’ 도메인의 이벤트가 시간과 장소의 연속성을 보이며 보도됨을 알 수 있었다.

이를 바탕으로 2012년 6월부터 2012년 12월 사이에 보도된 ‘보건/의료’ 분야 592,734건의 뉴스를 대상으로 전수 조사를 실시하였다. 조사의 편의를 위해 문서 내에 나타난 질병명, 시간 정보, 장소 정보는 한국전자통신연구원(ETRI)과 목원대에서 개발한 언어분석[11] 및 이벤트 정보추출기[12]를 활용해 자동으로 추출한 결과를 보정하지 않고 사용하였다. <표 1>은 전수 조사 결과이다.

<표 1> 보건/의료 뉴스 조사 결과

구분	뉴스 건수
보건/의료 뉴스 전체	592,734
시간 정보가 표현된 뉴스	314,421 (53%)
장소 정보가 표현된 뉴스	32,142 (5%)
질병명이 나타난 뉴스	115,215 (19.4%)
44개 연속성 질병명이 나타난 뉴스	2,142 (3.6%)
시간/장소/연속성 질병명이 함께 나타난 뉴스	175

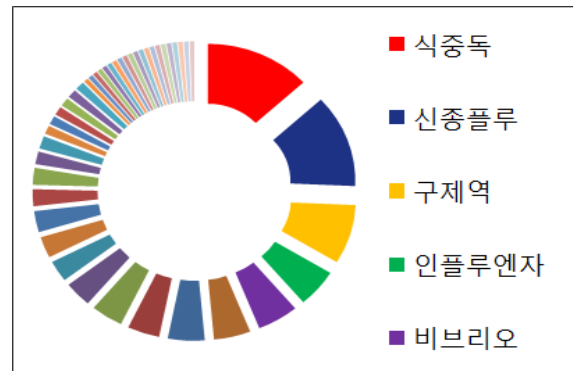
<표 1>에서 나타나듯이, 뉴스의 특성상 특정 사건 즉 이벤트의 발생 시간 정보를 포함하는 경

우가 많았고 (전체 592,734건 중 314,421건으로 53%), 사건 발생 장소 정보가 포함되는 경우는 상대적으로 적게 나타났다 (전체 중 5%). 또한 ‘보건/의료’ 분야의 뉴스임에도 불구하고 실제 명확한 질병명이 나타난 뉴스는 전체 19.4%에 그쳤음을 알 수 있었다.

앞서 설명한 바와 같이, ‘보건/의료’ 분야에 출현한 다양한 이벤트 중 본 논문에서는 ‘질병’ 도메인을 주요 이슈 이벤트로 선정하였으며, 그 중에서도 연속성을 갖는 이슈 이벤트를 선정하기 위해 아래와 같은 기준을 적용하였다.

- 1) 사회적 파급력이 있는지 (발생빈도)
- 2) 전염성(유행성)이 있는지
- 3) 초기 발생 이후 지속성이 있는지

조사 결과, 이러한 기준에 해당하는 질병명은 ‘구제역’, ‘식중독’, ‘신종플루’ 등 44개가 초기 선정되었으며, 해당 질병명이 나타난 뉴스 건수는 2,142건으로 전체 3.6%에 해당한 것으로 분석되었다. 연속성 분석을 위해서는 시간 및 장소 정보가 모두 표현된 뉴스만을 대상으로 하는데, 이런 조건을 만족하는 뉴스는 이 중에서 175건이 수집되었다.



<그림 1> 질병별 뉴스 분포

<그림 1>은 44개 질병 이벤트 후보 도메인의 분포를 도식화 한 것이다. 상위 질병을 보다 세부적으로 살펴보면, ‘식중독’, ‘신종플루’, ‘구제역’, ‘비브리오’, ‘인플루엔자/AI’ 순으로 나타났으며, 세부 질병별 뉴스 세부 건수는 <표 2>와 같다.

본 논문에서 제안하는 이슈 이벤트의 연속성을 분석하기 위해서는 소셜미디어에 보도하고 있는

이벤트가 앞서 제시한 바와 같이 1) 사회적 파급력과 2) 전염성, 3) 지속성이란 특성을 만족하며, 동시에 미디어 내의 문장에 시간과 장소정보가 명시적으로 표현되어 있는 문서를 대상으로 선정해야 한다. <표 2>의 상위 질병 이벤트의 연속성 적합 조건을 분석한 결과, <표 3>과 같은 특성을 발견하였다.

<표 2> 세부 질병별 뉴스 발생 건수

질병명	건수	질병명	건수
식중독	23	장염	2
신종플루	22	A형간염	1
구제역	14	감기	1
비브리오패혈증	9	녹내장	1
인플루엔자	9	눈병	1
AI	8	대사증후군	1
대장균	8	장출혈	1
독감	7	말기코환자	1
일본뇌염	7	말라리아	1
폐질환	6	무좀	1
c형간염	5	박테리아	1
폭염	5	백일해	1
수족구병	5	브루셀라	1
RS바이러스	4	비만	1
요로결석	4	성조숙증	1
결핵	3	수두	1
노로바이러스	3	신증후군	1
발열성질환	2	쓰쓰가무시	1
백내장	2	위염	1
보툴리눔독소증	2	유행성 눈병	1
비염	2	의료사고	1

<표 3> 이슈 이벤트 후보 도메인별 특성

질병명	연속성 적합 조건			적합여부
	전염/유행성	지속성	시간/장소	
식중독	0	0	0	0
신종플루	0	0	해외지명	-
구제역	0	0	0	0
비브리오패혈증	-	0	0	-
인플루엔자	0	0	해외지명	-
대장균	0	-	0	-

상위 질병 이벤트 중 ‘비브리오패혈증’은 단순 질병 발생의 경우로, 전염성이 없었고, ‘대장균’ 역시 지역이나 시간의 흐름에 따른 지속성이 발견되지 않아 대상 후보에서 제외되었다. 연속성 적합 조건을 모두 만족하는 후보 도메인은 ‘식중독/신종플루/구제역/인플루엔자’로 선정되었으나, 이 중에서 ‘신종플루’와 ‘인플루엔자’와 관련된 뉴스가 대부분 해외에서 동시다발적으로, 혹은 산발적으로 발병한 사례를 보도한 경우로, 장소의 연속성을

파악하기 어려운 점을 고려, 제외하였다. 본 논문에서는 이 같은 특성을 종합적으로 고려한 결과, 최종적으로 ‘식중독’과 ‘구제역’ 질병을 이슈 이벤트 후보로 정의, 이들 간에 시간과 장소의 흐름에 따라 어떠한 연속성이 분석되는지를 파악하기로 한다.

4. 이슈 이벤트 연속성 분석 시스템

이슈 이벤트 연속성 분석을 위해서는 소셜미디어에 나타난 텍스트 문장에서 이벤트 정보를 추출해야 한다. 이를 위해 본 논문에서는 한국전자통신연구원에서 개발한 이벤트 정보추출기[11]를 활용하였다. 이슈 이벤트 정보 추출 과정은 다음과 같다. 먼저 각 문장의 언어분석 결과를 바탕으로 이벤트가 출현할 가능성이 있는 ‘이벤트 문장’을 선별한다[12]. 이벤트 문장은 이벤트 정보의 핵심 요소로 사건의 구체적인 내용을 포함하고 있는 문장으로, 본 논문에서는 ‘구제역’과 ‘식중독’ 질병이 나타난 문장을 대상으로 삼았다. 이벤트 ‘발생 시간’과 이벤트 ‘발생 장소’는 연속성을 분석하기 위한 기준으로 사용된다.

평택 돼지농장 구제역 확진
 | 기사입력 2011-01-08 11:19 | 최종수정 2011-01-08 13:37

[1] <LOC:경기도> 내 <EVENT:구제역>의 마지막 보루였던 <LOC:평택시>가 결국 뚫렸다.

[2] <LOC:경기도> <EVENT:구제역> 재난안전대책본부는 <DATE:지난 7일> 의심 신고된 <LOC:평택시 고덕면 문곡리> 돼지농장의 돼지가 <EVE:구제역>으로 확진됐다고 <DATE:8일> 밝혔다.

[3] 이 농장은 <DATE:7일> 오전 어미돼지 1마리가 콧등에 물집이 생기는 등 <EVENT:구제역> 의심증상을 보이고 새끼돼지 20여 마리가 폐사했다고 신고했다.

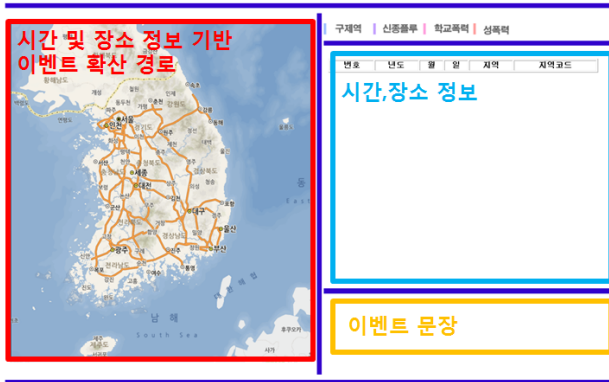
[4] 이에 따라 도 <EVENT:구제역> 재난안전대책본부는 이 농장 돼지 4천여 마리와 농장 주인이 <LOC:오성면 양교리>에서 키우는 돼지 2천480여 마리를 예방적 차원에서 살처분한 바 있다.

.....

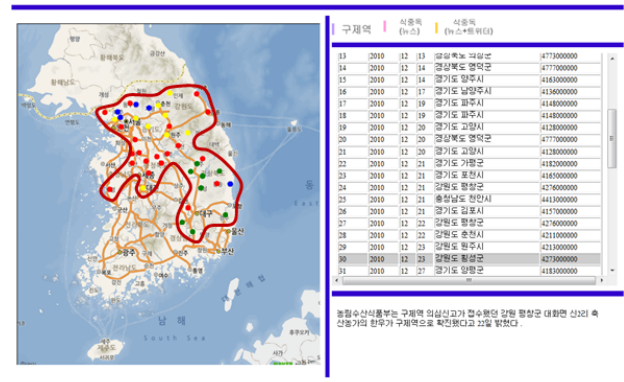
<그림 2> 뉴스미디어 언어분석 예



<그림 3> 이벤트 정보 추출 결과



a) 인터페이스 구성도



b) 실제 분석 결과 제시 예

<그림 4> 이벤트 연속성 분석 사용자 인터페이스

<그림 2>는 2011년 1월 8일의 보도된 ‘구제역’ 관련 뉴스의 자동 분석 결과를 예시한 것으로, 문장 내의 <EVENT>, <LOC>, <DATE> 태그는 각각 이슈 후보 이벤트, 이벤트 발생 시간 정보, 이벤트 발생 장소 정보의 후보를 의미한다[12]. 예시된 4 문장 중, 대상 이벤트와 시간, 장소 정보가 모두 포함된 문장은 2번째 문장으로, 최종 추출된 이벤트 정보를 요약하면 <그림 3>과 같다. 특히 실제 문장에 다양한 방식으로 표현된 시간 정보와 장소정보는 각각 8자리 숫자 (YYYY-MM-DD)와 GPS 좌표 체계로 정규화 된다[9].

<그림 4>는 이벤트 연속성을 시각적으로 분석하기 위한 사용자 인터페이스(User Interface)이다. <그림 4.a> 인터페이스 구성 요소로는 실험 문서에서 추출한 이벤트 발생 시간과 이벤트 발생 장소 정보를 구체적으로 나타내는 부분과 추출된 시간 및 장소정보를 기반으로 지도상에 이벤트 확산경로를 표시하기 위한 부분, 그리고 실험 문서에서 이벤트 정보가 추출된 실제 이벤트 문장을 확인하는 부분으로 구성되어 있다. <그림 4.b>는 실제 이벤트 정보가 분석된 결과를 제시한 예이다.

5. 이슈 이벤트 연속성 분석 결과

이벤트 연속성을 분석하기 위하여 분석 대상 도메인의 이벤트 정보를 수집하고 시간 및 장소

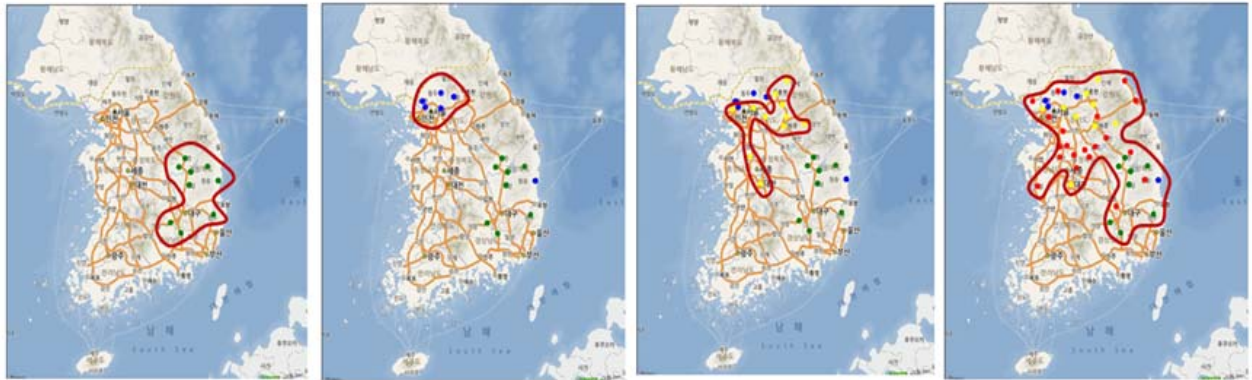
정보를 기준으로 이벤트의 확산 방향을 시각적으로 분석하였다. 분석 내용은 다음 두 가지 이다. 첫째, 연속성 분석 대상 도메인의 이벤트 발생과 이벤트 확산 유형을 분석한다. 둘째, 웹 뉴스와 소셜 마이크로블로그 간의 상관관계 및 두 소셜 미디어 간의 정보 확산 방향을 분석한다.

5.1 이슈 이벤트 연속성 분석 결과

이슈 이벤트 연속성 분석을 위해 ‘구제역’ 사건이 집중 보도된 2010년 11월부터 2011년 2월 사이의 신문기사 250건과 ‘식중독’이 주로 발생한 여름철을 대상으로 2012년 6월부터 2012년 11월 사이의 뉴스 71건을 수집하였다. <그림 5>는 ‘구제역’ 관련 대상문서를 이벤트 정보추출기[12]의 입력으로 활용, 실제 추출된 이벤트 정보를 시공간 축으로 분석 결과이다. 구제역이 발생한 시간과 지역을 기준으로 4단계로 분류를 하였다.

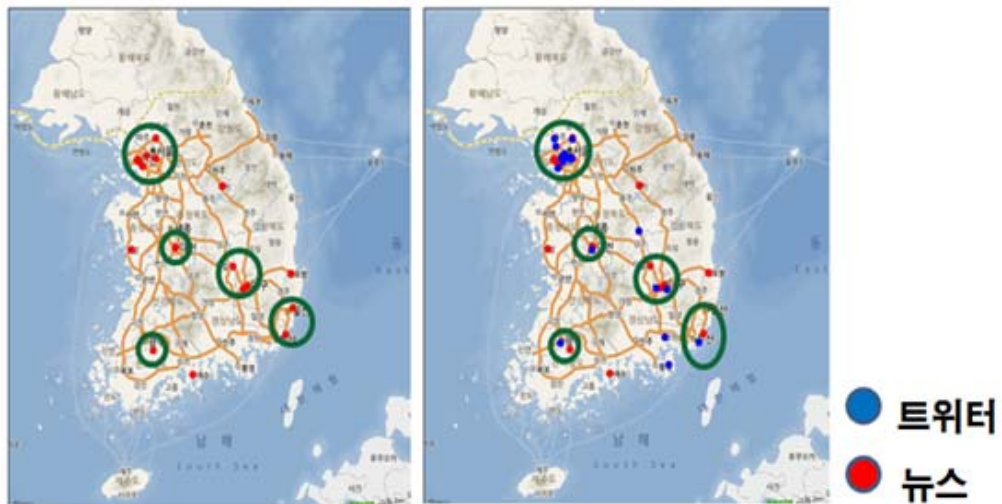
<그림 5.a>는 구제역의 1차 확산 단계로 2010년 11월 29일에서 2010년 12월 9일 사이에 발생한 이벤트들을 도식화한 그림이다. 경상북도 안동에서 최초로 발생하여 경상북도 예천군, 의성군, 영양군, 고령군 등 경상북도 내에서 지리적으로 가까운 곳에 위치해 있는 지역들만 이벤트 연속성을 보이고 있다.

<그림 5.b>은 구제역 2차 확산 단계로 2010년 12월 11일에서 2010년 12월 20일 사이에 발생한 이벤트들의 집합이다. 경기도 양주시, 과천시, 고양시 등 구제역 1차 확산 단계와 비슷한 현상으



a) 1차 확산 b) 2차 확산 c) 3차 확산 d) 최종 확산 결과

<그림 5> ‘구제역’ 이벤트 연속성 분석 결과



a) 뉴스 이슈 이벤트 발생 b) 트윗 이슈 이벤트 발생

<그림 6> 이슈 이벤트와 소셜미디어와의 상관 관계

로 경기도 지역 내에서 지리적으로 가까운 곳에 위치해 있는 지역들만 이벤트 연속성을 보이고 있다.

<그림 5.c>는 구제역 3차 확산 단계로 2010년 12월 21일에서 2010년 12월 31일 사이에 발생한 이벤트들의 집합이다. 구제역 3차 확산 단계는 전국적 확산 초기 단계로도 볼 수 있으며 경기도 지역 일대에서 강원도 지역 또는 충청북도 지역으로 이벤트의 연속성이 조금씩 확산되는 현상을 확인할 수 있다.

<그림 5.d>은 구제역 4차 확산 단계로 2011년 1월 1일에서 2011년 1월 15일 기간 사이의 발생한 이벤트 집합이다. 구제역 4차 확산 단계는 구

제역이 전국적으로 확산되었음을 확인할 수 있다. 정리해보면, 초기 구제역 1차 확산과 2차 확산 단계에서는 경상북도 지역과 경기도 지역 내에서 지리적으로 가까운 곳만 이벤트 연속성을 보였으나 3차 확산 단계 이후에는 전국적으로 이벤트가 발생하는 현상을 보였다.

5.2 이슈 이벤트와 미디어 간의 상관관계 분석 결과

이슈 이벤트와 소셜미디어 간의 상관관계를 분석하기 위해 웹 뉴스 미디어와 트위터를 대상으로 이슈 이벤트를 분석하였다.

<그림 6>은 웹 뉴스를 대상으로 ‘식중독’ 도메

인의 문서를 수집하고 해당 이벤트들의 연속성을 분석한 결과이다. <그림 6.a>의 식중독 발생 분포 분석 결과를 살펴보면, 2012년 6월 1일에서 2012년 11월 30일까지 발생한 이벤트의 집합이며 구제역 도메인과는 상반되게 지역적 연속성은 나타나지 않고, 시간의 흐름에 따라 주요 대도시에 질병이 밀집하는 현상만 보였다. 이는 ‘식중독’ 정보가 각 주요 대도시별로 관리, 뉴스에 보도 되고 있음을 암시한다.

웹 뉴스와 소셜미디어 간의 상관관계를 분석하기 위하여 ‘식중독이 발생했다’ 혹은 ‘식중독에 걸렸다’라는 언급을 한 트윗 문서를 수집하였다. 문서 수집 기간은 웹 뉴스에서 식중독 문서를 수집한 기간과 동일하게 2012년 6월 1일에서 2012년 11월 30일까지의 문서를 수집하였으며, 수집된 문서 중 GPS(Global Positioning System)정보 사용을 허용하는 사용자의 25개의 트윗 문서만 활용하였다. <그림 6.b>는 웹 뉴스와 소셜 웹 서비스간의 상관관계를 나타낸다.

동일한 기간에 수집된 웹 뉴스와 소셜미디어 모두에서 식중독이 비슷한 지역에서 발생한 것을 확인할 수 있다. 이는 사회적 현상이 뉴스와 소셜 개인 미디어 모두에 표현됨을 의미하며, 두 미디어간의 상관도가 매우 높음을 반증하는 예이다.

6. 결론

본 논문에서는 시간과 장보 축을 기반으로 이벤트들의 연속성을 분석하였다. 특히 사회적으로 파급력이 큰 사건/사고를 ‘이슈 이벤트’라고 정의, 도메인별로 발생하는 연속성 이벤트 유형과 특성을 규명하였다. 또한, 소셜미디어 상에서 발생한 연속성 이벤트를 자동으로 탐지, 이를 추적하는 이슈 이벤트 연속성 분석 시스템에 대해 기술하였으며, 이를 통해 추출된 특정 도메인에 대한 이슈 이벤트들의 연속성을 시간 및 장소 축으로 도식화하여 분석하고, 나아가 이슈 이벤트의 전과양상을 통해 소셜미디어간의 상관관계를 파악하였다.

이벤트 연속성 분석 결과, ‘구제역’ 도메인에서는 이벤트의 발생이 시간의 흐름에 따라 근접 지역으로 이벤트가 확산되다가 마지막에는 전국적

으로 확산되는 전과양상을 확인할 수 있었다. ‘식중독’ 도메인에서는 시간의 흐름에 따라 지역적인 이벤트 연속성은 나타나지 않았지만 소셜미디어인 뉴스와 트위터 간의 이벤트가 발생한 지역이 유사한 상관관계가 있음을 확인할 수 있었다.

향후 계획으로는, 이슈 이벤트 연속성을 분석하기 위해서는 근본적으로 소셜미디어 문서에서 자동으로 이벤트 문장을 추출 성능에 따라 연속성 분석 성능 역시 관련되어 있으므로, 이벤트 자동 추출 성능 향상을 위한 다양한 어휘 자질을 탐구하고자 한다. 또한, 다양한 도메인으로 연속성 분석을 확장, 도메인별 특성을 파악하고자 한다.

참 고 문 헌

- [1] 허윤, 이상우 (2012), 다매체 환경에서의 미디어 간 경쟁과 대체, **한국언론학보**, 56(4), 29-54.
- [2] <http://word.tta.or.kr/terms/terms.jsp>
- [3] TDT (2004). Annotation Manual Version 1.2 (www.itl.nist.gov/iad/mig//tests/tdt/)
- [4] Allan, J. Wade, C., & Bolivar, A. (2003). Retrieval and novelty detection at the sentence level. *The proceedings of SIGIR 2003*, 314 - 321.
- [5] Swan, R., & Allan, J. (2000) Automatic generation of overview timelines. *The proceedings of SIGIR 2000*, 49 - 56.
- [6] He, Q., Chang, K., & Lim, E.-P. (2007). Analyzing feature trajectories for event detection. *The proceedings of SIGIR 2007*, 207 - 214.
- [7] Haghani P, Michel S., & Aberer K. (2011) Tracking hot-k items over Web 2.0 streams. *Datenbank systeme in Business, Technologie und Web*, 14, 105 - 122.
- [8] Mislove, A., Lehmann, S., Ahn, Y., Onnela, J., & Rosenquist, J. (2010). Pulse of the nation: Us mood throughout the day inferred from twitter, *Accessed November 22*.
- [9] Sakaki, T., Okazaki, M., & Matsuo, Y (2010). Earthquake shakes twitter users:

real-time event detection by social sensors. *The Proceedings of the 19th international conference on World wide web (WWW '10)*. 851~860.

- [10] Imran, M., Elbassuoni, S., Castillo, C., & Meier, P. (2013). Practical extraction of disaster-relevant information from soial media. *The Proceedings of the 22nd international conference on World wide web (WWW '13)*. 1021-1024.
- [11] Oh H-J, Lee CK, Lee CH (2012). "Analysis of the Empirical of Contextual Matching Advertising for Online News", *ETRI Journal*, 34(2), 292-295.
- [12] 윤보현 (2012) 문장레벨 이벤트 분류를 위한 알고리즘 비교, **한국정보기술학회논문지**, 10(6), 171-176.



윤 보 현

1999 고려대학교 컴퓨터학과
(이학박사)

1999~2002 : 한국전자통신연구원
선임연구원(팀장)

2003~현재 : 목원대학교 컴퓨터교육과 부교수
관심분야: 자연어처리, 정보검색, 시맨틱웹,
e-Learning

E-Mail: ybh@mokwon.ac.kr



오 호 정

2000 충남대학교
컴퓨터과학과(이학석사)
2008 한국과학기술원
컴퓨터공학과(공학박사)

2000~현재 한국전자통신연구원 지식마이닝연구
실 선임연구원

관심분야: 정보검색, 질의응답, 소셜마이닝, 빅데
이터 정보처리

E-Mail: ohj@etri.re.kr



김 현 기

2005 Universify of Florida
전산학(박사)

1995~현재 : 한국전자통신연구원
지식마이닝연구실장

2012~현재 : 과학기술연합대학원대학교 컴퓨터소
프트웨어공학 겸임교수

관심분야: 자연어처리, 정보검색, 질의응답, 소셜
빅데이터 분석

E-Mail: hkk@etri.re.kr