

도메인 지식 기반 이슈 탐지 모델링

- 해외 발생 감염병 국내 유입 이슈를 중심으로

Socio-National Issues Detection Modeling based on Domain Knowledge

- Focusing on the Issue of Increase in Domestic Inflow Infectious Diseases

황미녕, 이승우

한국과학기술정보연구원 융합기술연구본부

Mi-Nyeong Hwang(mnhwang@kisti.re.kr), Seungwoo Lee(swlee@kisti.re.kr)

요약

빅데이터 관련 기술의 발전으로 공공 보건 분야 등을 필두로 데이터에 기반한 정책을 결정하는 체계적인 방법론에 대한 관심이 증가하고 있다. 본 연구는 데이터를 기반으로 국가, 사회의 주요 이슈를 지능적으로 탐지하기 위해서 도메인 전문가와의 협업을 통해 이슈 탐지 모델을 개발하는 방법을 제안한다. 우선, '해외 발생 감염병 국내 유입' 이슈를 대상으로 이슈에 영향을 주는 요인을 도출하고, 영향 요인을 대표하는 변수들을 설정한다. 다음으로 시스템 다이내믹스 기법을 이용하여 각 영향요인 간의 인과 분석을 통해 인과지도 구성하여 영향력 높은 주요 요인들을 찾아낸다. 이 과정에서는 데이터 분석가와 감염병 도메인의 전문가와의 협업을 통해 실증적인 모델링을 진행한다. 이러한 도메인 지식 기반 이슈 탐지 모델을 기반으로 하여 상시 모니터링이 가능한 이슈 탐지 체계가 구축되면 더욱 효과적인 정책 의사 결정이 가능할 것이다.

■ 중심어 : | 이슈 탐지 | 이슈 모니터링 | 모델링 | 시스템 다이내믹스 | 감염병 탐지 |

Abstract

As the big data technologies advance, there is an increasing interest in systematic methodologies for data-based policy determination especially in the public health area. This study proposes a method to develop an issue detection model through the collaboration with domain experts in order to intelligently detect major socio-national issues on infectious diseases based on data. At first, the factors influencing the 'domestic inflow of foreign infectious diseases' are determined and variables representing the factors are set. Thereafter, by using system dynamics methods, the causal analysis is made to find causal map indicating main influential factors. In this process, an empirical modeling is conducted through collaboration between data analysts and experts in the infectious disease domain. The proposed issue detection approach based on domain knowledges will make it possible to make a decision on policies more efficiently if the detection system is capable of continuous monitoring of the related issues.

■ keyword : | Horizon Scanning | Issue Monitoring | Modeling | System Dynamics | Infectious Disease Detection |

* 본 연구는 2017년도 한국과학기술정보연구원(KISTI) 주요사업 과제로 수행한 것입니다

접수일자 : 2017년 10월 30일

심사완료일 : 2017년 12월 13일

수정일자 : 2017년 12월 07일

교신저자 : 이승우, e-mail : swlee@kisti.re.kr

1. 서론

미래연구는 전통적으로 전문가 그룹의 통찰과 직관에 중점을 두던 미래예측(Forecasting)에서 대응 방안을 포함하는 미래전략(Strategic foresight)으로 발전 중이다. 이는 과거와 달리 기술의 빠른 발전과 세계화, 복잡성, 불확실성이 높은 사회에서 데이터에 기반한 체계적인 분석에 의한 미래예측과 대응 전략의 수립이 요구되고 있기 때문이다[1]. 특히, 정량적 데이터 분석에 의한 미래연구는 최근에 주목받기 시작하고 있으며 빅데이터에 대한 기술이 발달함에 따라 점차 비중이 커지고 있다.

2000년대 초반부터 영국과 미국 정부를 중심으로 ‘개방 정부와 개방 데이터(Open Government and Open Data)’ 전략에 따라 다양한 공공 데이터가 공개되었다. 이후 이러한 공공 데이터를 가공, 분석하여 새로운 가치를 창출하는 데이터 가공 산업이 활성화 되었다[2]. 공공 데이터뿐만 아니라 인터넷과 스마트폰의 확산으로 개인이 생산해내는 SNS 데이터는 실시간성을 가지는 사회 현안이나 이슈에 대한 즉각적인 정보로서 활용이 가능하다. 주요 선진 국가들은 각종 현안 해결뿐만 아니라, 미래예측 및 전략적 대응에서도 빅데이터를 활용하기 시작하였고, 데이터 분석에 기반한 호라이즌 스캐닝을 통해 미래 잠재적인 위협에 대한 징후를 포착하여 선제적 대응방안 모색 중이다[3].

그럼에도 불구하고 문제는 남아있다. 획득 가능한 데이터는 비록 급격히 늘었지만, 정책결정자들이 실제 사회 문제를 해결하기 위해 적절한 형태의 정보를 제한된 시간 이내에 얻는 것은 여전히 어려운 실정이다. 예를 들어, 구제역(foot-and-mouth disease)이나 메르스(MERS :Middle East Respiratory Syndrome)와 같은 질병을 효과적으로 관리하거나 국지성 폭우로 인한 침수나 산사태 등의 재난 상황에 효율적으로 대처하기 위한 정책 결정은 데이터의 신속한 처리를 통해 제한된 시간 안에 이루어져야 한다. 그러나 현재 이러한 문제를 잘 해결할 수 있는 효과적인 빅데이터 처리 도구 및 환경은 확립되지 않은 상태이다[4]. 특히, 2015년의 한국의 메르스 사태에서 경험했듯이 메르스 중앙방역대

책반을 2년 전부터 운영을 하면서도 사람 간의 지속적인 전파에 대한 증거가 없었다는 WHO의 회의 권고문을 듣고, 해외발생 감염병에 대한 정보 부족 및 안일한 대응으로 전국이 사회경제적 혼란에 빠지기도 했다[5]. 이는 해외발생 신종 감염병의 국내 유입에 대한 대응에서 주요한 요인을 선별하고 선별된 주요 요인들에 대한 대비 및 대응을 하는 것이 중요하다는 의미이다.

해외발생 감염병이란 국내 발생은 거의 없으나 사람을 포함한 매개물을 통해 국외에서 국내로 유입 또는 전파되는 감염병을 의미하여 메르스나, 지카 바이러스의 예를 들 수 있다[6]. 해외발생 감염병은 조기감지와 빠른 대처를 통해 발생지역 내외로 확산을 방지하는 것이 중요하다. 전세계적인 네트워크를 보유하고 있는 국제기구 및 선진국들은 이미 다양한 정보 모니터링과 감시, 정보 공유를 통해 감염병 발생 초기 단계에서 감지 및 대응하기 위한 활동을 수행하고 있다. 이런 보건당국 및 의료기관이 제공하는 정형화된 지표 기반의 질병 감시체계는 공식발표까지 필요한 행정처리 및 적시성, 신속성 관점에서 한계가 대두되었고, 이를 보완하기 위해 SNS 및 지역 뉴스기사 등에서 언급되는 질병 관련 정보를 실시간 수집, 분석하여 더 효과적으로 신속한 질병 감시 체계를 만들기 위해 노력해왔다[7][8].

감염병의 발생 곡선에서 발생 단계의 가장 빠른 정보 획득 소스를 SNS로 예측한 연구가 있었다. 감염병이 확산기에 들어서는 시점까지 감염병 정보를 획득할 수 있는 소스는 소셜 네트워크 및 블로그와 같은 개인 간 커뮤니케이션 수단이며 시간이 지날수록 정보 획득 소스의 공개범위가 넓어진다[9]. 이런 기존 연구에서 보듯이 감염병 발생을 탐지하는 것은 공공 기관의 공식적인 발표 이전에 지역 신문 기사, 소셜 네트워크 데이터 등의 비정형 데이터를 복합적으로 분석해야할 필요성이 대두된다[6]. 현재 WHO 등의 공공기관의 발표, 뉴스, SNS 등의 비정형 텍스트를 분석하여 감염병 발생을 모니터링 하는 서비스로는 Global Public Health Intelligence Network(GPHIN)¹, BioCaster, HealthMap², Medical Information System(MediSys)³, Pattern-based

1 GPHIN: <https://gphin.canada.ca/>

2 HealthMap: <http://www.healthmap.org/>

3 MediSys: <http://medisys.newsbrief.eu/>

Understanding and Learning System(PULS)⁴, Program for Monitoring Emerging Infectious Disease (ProMED-Mail)⁵ 등이 있으며, 이들은 기본적인 주요 감염병 키워드를 기반으로 다국적 뉴스 정보를 통합하여 제공한다[10-12].

본 연구에서는 국가, 사회를 위협하는 주요 이슈를 지능형으로 탐지하는 기술을 해외발생 감염병의 국내 유입 이슈를 대상으로 하여 이슈에 영향을 주는 주요 영향요인을 도출하고 요인간의 인과분석을 통해 이슈 탐지 모델링을 하는 방법에 대해 기술하고자 한다. 이런 과정에서는 빅데이터 분석이 위주의 모델 수립이 아닌 감염병 도메인의 연구자들과의 협업 과정을 통해 진행되었고 이 모델링 과정에 대하여 구체적으로 다룰 것이다.

2. 관련연구

다양한 영역의 데이터를 살펴 미래에 문제를 발생시킬 가능성이 있는 요소를 감지하는 것을 호라이즌 스캐닝(Horizon Scanning)이라 부른다[13]. 망망대해의 수평선에서 잠재적 위협이 될 수 있는 요소가 돌출하는 것을 감지하여 탐지하는 것과 같은 것이다. 이미 영국과 싱가포르, 캐나다 등의 나라에서는 효과적인 미래예측을 위해 국가 차원의 호라이즌 스캐닝 전담기관을 설치하여 운영하고 있다[14][15]. 정부 주도의 전담 기관의 설치가 아니어도 다양한 데이터를 기반으로 하여 주요 이슈를 발굴하고 미래예측을 하는 연구는 빅데이터 관련 기술의 발달과 병행하여 진행되고 있다[3].

2.1 정책 지원을 위한 시스템 다이내믹스의 활용(영국)

영국 정부는 미래 예측을 위해 최신 과학이론과 증거 기반의 의사결정 수립 방안을 마련해주는 Horizon Scanning Programme을 운영하고 있으며 현재는 데이터 분석 팀과 통합하여 Horizon Scanning Programme의 시스템을 개발, 운영하고 있다[16]. 영국의 Horizon Scanning 프로젝트는 10년에서 15년 후의 중장기 미래 사회를 예측하는 것으로 2020년대의 성장 기회를 찾기

위한 기술 및 혁신에 대한 보고서, 세계 경제 보고서, 미래 에너지 보고서 등이 있다. 현재 진행 중인 미래 예측 프로젝트로는 도시의 장기적인 미래 모습과 영국 사회의 고령화로 인한 미래 사회가 있다.

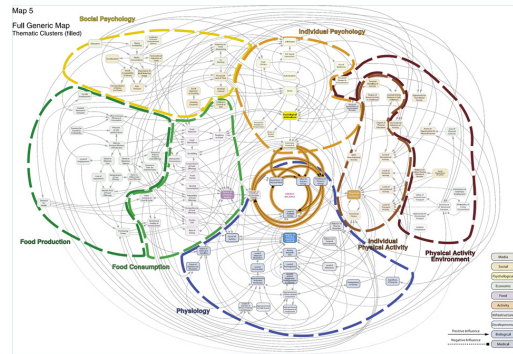


그림 1. 영국 비만대책 미래예측의 인과지도

기존 연구로는 영국 사회의 40년 후를 예상하는 비만 대책 프로젝트 Tackling Obesities: Future choice가 있다. 영국인의 비만 문제가 사회를 위협하는 중요한 사항으로 판단하고, 과학적 근거를 기반으로 비만대책을 수립하였다. 비만에 영향을 주는 주요 요인들을 미디어, 사회적, 심리적, 경제적, 영양적, 활동성, 사회구조, 발달학적, 생물학적, 의학적 카테고리로 구분하여 도출하고, 인과관계 지도를 [그림 1]과 같이 구성하였다. 이러한 사회 시스템의 복잡한 피드백 시스템을 연구하는 시스템 다이내믹스 기법을 이용하여 모델을 만들고 4개의 시나리오를 만들어서 영국 국민의 비만 방지 정책 수립 부서에 반영하였다[17][18]. 미래에 당면할 사회 문제를 동태성에 중심을 두는 시스템 다이내믹스 기법으로 300여명의 분야별 전문가들이 협업하여 인과지도를 만들어낸 체계적인 시도였다는 것에 의의를 둘만 하다.

2.2 H5N1 위험 분석 사례(싱가포르)

싱가포르 정부에서는 주변국에서 발생하는 조류 인플루엔자 바이러스(H5N1)의 국내 유입 위험을 체계적으로 대비하기 위한 노력을 일찍이 진행해왔다. 감염병 전문가 그룹과 정부 정책 결정자, RAHS(Risk Assessment and Horizon Scanning Programme Office)의 데이터

4 PULS: <http://puls.cs.helsinki.fi/>

5 ProMED-mail: <https://www.promedmail.org/>

분석가들의 협의체를 구성하여 H5N1의 싱가포르 유입에 영향을 줄 수 있는 주요한 위험 요인에 대해 Delphi 기법으로 분석하였다[19]. 도출된 위험 요인에 대해서 과 분석을 통해 인과지도를 구성하고 이를 기반으로 여러 시나리오를 구성하여 정책 결정 부서에 지원하였다. 이 연구에서는 국가·사회의 안전을 위협하는 신종 감염병에 대해 관,학,연의 협업 과정을 통해 다방면의 위험 요인을 도출해내고 인과 분석을 통해 형태학적인 모델링을 진행한 것에 의의가 있다. 그러나, 질병의 특성, 질병의 동태성, 인구 이동 등의 여러 요인에 따라 해당 모델을 범용적으로 적용하기 어렵다는 한계를 가진다. 또한, 이 모델은 싱가포르 국가의 지리적, 사회적 특성에 기반하고 있기에 환경이 다른 한국 상황에 그대로 반영하기에는 적절치가 않다.

3.3 국가·사회주요 이슈에 대한 이슈-데이터 맵 구축 사례(KISTI)

국가·사회 주요 이슈를 발굴하고, 도메인 전문가와의 협업을 통해 이슈를 해결하는데 필요한 이슈-데이터 맵을 구축하는 연구도 진행되어 왔다[20]. 이슈-데이터 맵의 구축 목적은 이슈 관련 데이터, 연구자, 의사결정 체계를 쉽게 찾고 활용할 수 있도록 함으로써 이슈에 대한 정책과 R&D를 연결하는 통로 역할을 하고 근거 기반의 의사결정을 지원하기 위함이다. 이 연구에서는 국가·사회 주요 이슈 10종을 우선적으로 선정하고, 단순히 많은 정보를 수집하는 것이 아니라 미래의 변화에 영향을 미치는 요인을 중심으로 데이터를 수집하고 모니터링 하기 위해 STEPPER(Society, Technology, Environment, Population, Politics, Economy, Resources) 관점에서 영향요인을 도출하여 해당 이슈 도메인 전문가들의 검증 과정을 거쳤다. 각 영향요인 별로 관련된 세부 모니터링 항목을 분석하여 데이터 맵을 구축한다.

이러한 국가·사회 주요 이슈-데이터 맵의 구축은 신속한 정보의 수집과 이를 활용한 적절한 초기 대응을 지원할 수 있다. 하지만, 단순 이슈-데이터 맵의 구축에서 나아가 이 데이터를 기반으로 하는 활용 서비스의 제공이 되어야 이슈에 대한 탐지와 대응 체계를 발전시킬 수 있을 것으로 생각된다.

3. 본론

메르스는 2015년 국내에 유입되기 전 2012년에 중동 지역에 발병하였고, 2014년에 사우디아라비아에서 크게 유행하였으나 이에 대한 심각성 탐지에 실패하고 국내로의 유입에 사전 대비를 하지 못해 경제적, 사회적 피해가 극심하였다. 이런 국가·사회적 위기를 유발하는 잠재적 이슈를 신속하게 탐지하는 것은 다양한 데이터를 통합, 분석하여 이슈를 조기에 적극적으로 대응하기 위한 기반이 되는 기술이다.

지금까지의 이슈 탐지(Horizon Scanning) 기술은 텍스트 속에 포함된 키워드의 빈도수의 변화를 보고 이상 징후를 포착했다면 본 연구에는 도메인 지식 기반으로 해당 이슈에 영향을 주는 영향요인을 사전에 도출해내고 이 영향요인 간의 인과를 분석하여 인과지도를 구성해 이슈 탐지 모델을 생성한다. 이 모델을 기반으로 하여 주요 이슈의 이상 징후를 포착하기 위해서는 어떤 요인들을 중점적으로 모니터링할 것인지 판단이 가능해진다. 본 연구에서는 주요 이슈는 ‘해외발생 감염병의 국내 유입’로 한정하여 도메인 전문가와의 협업을 통해 이슈 탐지 모델링을 진행한 과정을 서술하기로 한다.

3.1 협업을 통한 모델링과정

도메인 지식 기반 이슈 탐지 모델링은 빅데이터 분석가, 감염병 도메인 전문가, 시스템 다이내믹스 전문가로 구성된 커뮤니티를 통한 토론과 협력을 통해 [그림 2]와 같이 진행되었다. 이 중 감염병 도메인 전문가 그룹은 미생물, 바이러스 전공의 교수 2명, 약학대학 교수 1명, 수의대학 교수 1명, 바이오인포매틱스 분야 연구원 1명의 5명으로 구성되었다.

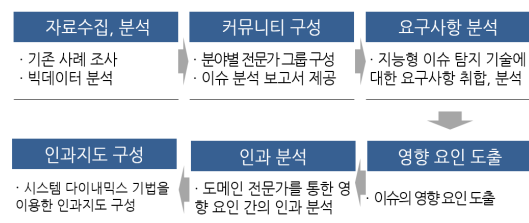


그림 2. 도메인 지식 기반 이슈 탐지 모델링 과정

표 1. '해외 발생 감염병의 국내 유입' 이슈의 영향요인 도출

구분	영향요인	변수
감염병/병원체 특성 (생태/환경)	전파력	전파경로
		임상증상 나타난 이후 전파 가능성
		무증상 및 경미한 증상 중 전파 가능성
		잠복기간 중 전파 가능성
		사람-사람 간 전파 가능성
		동물-사람 혹은 이종 동물 간의 전파 여부(가능성)
		바이러스 적응력 획득 여부
		잠복기
		기초감염재생산지수(RO: basic reproduction number)
	치사율(사망률)	치사율
발병률(morbidity rate)	발병률	
병원체의 변이 가능성	병원체 변이 가능성(다양성)	
계절/기후적 영향	병원체가 활동하기 적합한 국내 계절/기후 환경 정도	
매개체 특성 (생태/환경)	계절/기후적 영향	매개체가 활동하기 적합한 국내 계절/기후 환경 정도
	매개체 이동(유입)	매개체(예: 모기)의 국내 유입 여부
		매개체 서식지의 확산
	사람과의 접촉 가능성	매개체의 서식지 위치와 규모
매개체의 양식/사육 여부 및 환경		
예방, 치료 및 진단 특성 (기술)	효과적인 예방법	효과적인 백신 유무 백신 접종률
	효과적인 치료법	치료법
		치료제
	병원체 감염 역학 지도	병원체 감염 역학지도 확립 유무
	발병국의 효율적인 진단 기술 수준	진단 기술 유무
		진단 기술 보유 여부 진단 키트 보급률
국가(정책) 특성 (정책)	발병국의 출국자 관리 통제 능력	출국자 관리 통제 능력
	발병국의 보건 인프라 및 대처 능력	인프라(음압병동 시설 등)
		질병 관리 통제 능력(발병 지역 격리, 환자 격리, 접촉자 격리 등)
		투입 인력
	발병국의 국가(정책)적 노력 및 효과	예산
		정보 공개 수준
		신변중 감염병 치료, 진단, 예방 기수에 대한 신속허가 심의제도 유무 감염병 발병 경험 유무
	국내 입국자 관리 통제 능력	입국자 관리 통제 능력
국내 보건 인프라 및 대처 능력	질병 관리 통제 능력(발병 지역 격리, 환자 격리, 접촉자 격리 등)	
국내 감염병 예찰 프로그램 운영 여부	감염병 예찰 프로그램 운영 여부	
사회적 특성 (사회)	발병국과의 국제 교류 정도	국제 행사 개최
		왕래 인구 수
		항공/선박 왕래 수
	발병국의 사회(국민)의 인식 수준	감염병의 위험 정도와 예방법 등에 대한 인지 수준(국가의 특수한 문화와 보건 환경)
	감염병에 대한 국제 언론의 관심	감염병을 다루는 (국내) 언론 기사 건수
	감염병 발생 사건(이벤트) 수	해외 지역의 감염병 발생 환자 수
해외 지역의 감염병 발생 사례		

먼저 기존의 ‘해외발생 감염병의 국내유입’ 이슈에 대한 이슈-데이터 맵 구축 보고서를 기초 자료로 하여 싱가포르 RAHS의 조류 인플루엔자 위험 분석 사례 등을 조사하여 커뮤니티에 제공하여 상이한 전공을 가진 전문가들의 편향성을 제어하도록 노력하였다[19]. 또한 해당 이슈를 해결하기 위한 지능형 탐지 기술에 대한 감염병 도메인 전문가들의 요구사항과 미래연구 분석가들의 요구사항을 설문조사 및 심층 인터뷰(FGI:Focus Group Interview)를 통해 수렴, 분석하여 지능형 이슈 탐지 기술의 기능에 대해 구체적이고 명확하게 정의하였다. 그리고 오프라인 포럼을 통한 브레인스토밍(Brainstorming), 동일한 주제에 대해 집단의 사고의 폭을 넓혀가는 브레인라이팅(Brainwriting) 등을 통해 후보 요인들에 대해 의견을 나누고, 이를 바탕으로 하여 이슈에 영향을 주는 영향 요인(driving force, driver, factor)들을 도출해낸다. 여러 번의 반복된 도메인 전문가들의 피드백을 거쳐 영향요인 목록을 최종 결정한다. 이후 요인 간의 인과분석을 데이터 분석가와 도메인 전문가, 시스템 다이내믹스 전문가가 협업하여 진행하면서 인과지도를 구성하게 된다. 다음 절에서는 이슈에 영향을 주는 요인을 도출하고 인과분석을 거쳐 인과지도를 구성하는 과정에 대해 상세히 서술하고자 한다.

3.2 영향 요인 도출

‘해외 발생 감염병의 국내 유입’ 이슈에 대해 영향을 주는 요인들은 단순하지 않다. 낙타를 통해 감염된다고 생각한 메르스에 대한 정확한 정보가 부족하여 감염 대비가 늦었고, 실제 사람 매개를 통해 국내로 유입될 가능성에 대해 생각지 못하여 국내 의료진의 대응이 신속하지 못했다[5]. 반면에 지카 바이러스처럼 감염환자가 국내로 유입은 되지만 바이러스 확산될 국내 환경이 적합하지 않아서 감염병의 국내로의 확산이 되지 않는 경우도 있기에 신종 또는 재출현 감염병의 국내 유입에 영향을 주는 요인을 도출하는 것은 간단한 일은 아니다.

그리하여, 해외에서 발생하는 감염병의 국내 유입 가능성에 영향을 주는 요인들은 문헌 분석, 도메인 전문가의 그룹 토의 및 자문을 거쳐 [표 1]과 같이 생태/환경, 기술, 정책, 사회로 구분을 하여 요인을 도출하고 각

영향요인을 현실세계에서 구체적으로 표현할 수 있는 항목인 변수(interface variable)를 나열하였다.

A. 감염병/병원체 특성<생태/환경>

- a. 전파력: 감염병의 확산에는 어떤 전파경로로 감염이 되는지, 임상증상이 나타나기 전후로 전파 가능성이 있는지, 이종 간의 전파 가능성이 있는지, 병원체가 바이러스인 경우 바이러스 적응력 획득이 되었는지, 잠복기간, 감염병의 기초감염재생산지수(RO: basic reproduction number) 등이 변수가 된다.
- b. 치사율(사망률): 감염병의 치사율은 감염병의 발생 억제와 확산 저지를 막는 국가의 노력, 국제적인 협력에 영향을 미치는 요인이 된다.
- c. 발병률(morbidity rate): 발병률이 높을수록 감염병의 확산 및 이동의 가능성이 높다. 동일 질병에 대해서도 지역, 환경, 백신 보급률에 따른 발병률이 다르므로 지역별로 신뢰할 수 있는 기관의 발표 수치의 확보가 필요하다.
- d. 병원체의 변이 가능성: 특히, 병원체가 바이러스인 경우 변이 다양성이 높다. 이는 바이러스의 변이에 대한 지속적인 공공 DB의 구축 및 정보의 공유도 중요하다.
- e. 계절/기후적 영향: 병원체가 활동, 증식할 수 있는 기상 조건과 국내 환경이 이에 부합되는 정도를 의미한다. 이는 감염병에 따라 중요 여부가 다르기에, 감염병에 따라 가중치를 다르게 적용해야 할 필요가 있다.

B. 매개체 특성<생태/환경>

- a. 계절/기후적 영향: 모기, 진드기, 철새 등의 매개체가 활동, 증식할 수 있는 기상 조건과 국내 환경에 부합되는 정도의 요인을 의미한다.
- b. 매개체 이동(유입): 모기, 진드기 및 동물 등 매개체를 통해 감염되는 감염병의 경우 매개체의 발생과 이동은 감염병 발생 및 확산과 밀접한 관계가 있기에 매개체의 국내 유입 여부, 매개체 서식지의 확산 등의 변수로 나타낼 수 있다.

- c. 사람과의 접촉 가능성: 모기, 진드기, 조류, 쥐 등의 매개 감염병의 경우에는 인간과의 접촉 여부, 정도 등의 가능성에 따라 감염병의 확산이 되므로 매개체의 서식지 위치와 규모, 매개체의 양식/사육 여부 및 환경 정도 등이 변수가 된다.

C. 예방, 치료 및 진단 특성<기술>

- a. 효과적인 예방법: 해당 감염병에 대한 백신과 같은 효과적인 예방법이 있는 경우 감염병의 발생과 확산을 저지하고, 국내 유입의 가능성을 대비할 수 있는 요인으로 감염병 발생국 및 주변국, 국내에서의 백신 접종률이 구체적인 변수로 사용될 수 있다.
- b. 효과적인 치료법: 해당 감염병에 대한 치료제와 치료법이 있는 경우 효과적으로 질병에 대한 치료가 가능하므로 감염병 확산에 대한 대응 요인이 된다.
- c. 병원체 감염 역학 지도: 신, 변종 감염병 원인에 감염된 사람 혹은 동물 간의 전파 경로를 반영한 역학지도의 활용도는 주변국으로의 감염병 전파 차단 대책 수립에 필수적인 요소로 역학지도의 확립 유무가 변수로 표현이 된다.
- d. 발병국의 효율적인 진단 기술 수준: 감염병 조기 진단은 감염병 유입 시 빠른 대처 및 방역조치에 필수적인 요소이다. 발병국의 효율적인 진단 기술 수준은 감염병 발생, 확산을 조절할 수 있는 요인이 된다.

D. 국가(정책) 특성<정책>

- a. 발병국의 출국자 관리 통제 능력: 국가를 이동하는 해외여행객이 많을수록 감염병의 해외 유입 사례가 증가하는 경향을 보인다. 발병국의 출국자 관리 통제 능력이 국외로의 감염병 확산을 저지하는 요인으로 볼 수 있다. 또한, 발병국에서 감염병이 국내로 유입시 유입되는 사람에 대한 정보를 발병국에서 제공받을 수 있는 국가 간의 정보의 공유도 요구된다.
- b. 발병국의 보건 인프라 및 대처 능력: 발병국의 보

건 인프라 수준이 감염병 발생을 조절하고, 감염병 발생 시 대처 능력에 따라 감염병 확산을 조절할 수 있는 상시적인 국가 보건 인프라 수준을 의미한다.

- c. 발병국의 국가(정책)적 노력 및 효과: 감염병 발생 시 해당 발병국의 대응 능력으로 감염병 확산을 저지하기 위해 국가적인 노력의 수준으로 예산, 투입인력, 정보 공개 수준, 신변종 감염병 치료, 진단, 예방 기술에 대한 신속허가 심의제도 유무, 유사 감염병 발병 경험 유무 등의 구체적인 변수로 표현된다.
- d. 국내 입국자 관리 통제 능력: 국내 입국자의 직전 여행 국가 추적 능력, 입국자의 이상 징후 탐지 및 입국 후 이동지 추적, 통제 능력 등을 의미한다.
- e. 국내 보건 인프라 및 대처 능력: 감염병 발생을 대비, 대처할 수 있는 상시적인 국내의 보건 인프라 수준을 뜻한다.
- f. 국내 감염병 예찰 프로그램 운영 여부: 예찰(Surveillance)을 통해서 다양한 감염병의 사전 발생 가능성을 감시하는 것으로 예를 들어, 고병원성 조류 인플루엔자 검지를 위해 주기적으로 철새를 예찰하고, 미국의 웨스트나일 바이러스 유입 여부를 확인하기 위해서 일정 거리 별로 닭농장의 닭 혈청 검사를 진행하는 것이다.

E. 사회적 특성<사회>

- a. 발병국과의 국제 교류 정도: 감염병 발생국과의 교류 정도는 감염병 유입 가능성을 높이는 요인으로 볼 수 있다. 해당 발생 국가와의 인구 이동, 물류 교류, 주요 국제 행사 개최 등의 구체적인 변수를 들 수 있다.
- b. 발병국의 사회(국민)의 인식 수준: 발병국의 국민의 감염병 및 보건에 대한 인식 수준으로 조류 인플루엔자 위험성이 높은 지역에서의 가금류 시장의 위생 상태에 대한 인지 부족이나, 수인성 질병에 대한 정보가 부족하여 식수 사용에 대한 인지 부족 등 해당국 국민의 보건 의식 수준을 의미한다.

- c. 감염병에 대한 국제 언론의 관심: 위험성이 높은 감염병의 경우 국제기구의 언론 공표나 발병 사건 보고 등의 국제 언론 보도 횟수가 증가하게 되어 언론 기사 건수를 통해 국제적인 심각도를 알 수 있게 된다. 구체적으로 감염병을 다루는 국제 언론 기사 건수를 변수로 활용할 수 있다.
- d. 감염병 발생 사건(이벤트) 수: 구체적으로 해외에서 발생하는 감염병 발생 사례나 환자 수를 변수로 사용할 수 있다. 이 경우에는 다양한 정보 제공 소스를 활동하되 신뢰성 있는 정보를 선별하여 활용해야 한다.

이와 같이, 감염병 도메인 전문가와 협업하여 ‘해외발생 감염병의 국내 유입’ 이슈에 영향을 미치는 요인들을 환경, 기술, 정책, 사회 영역 관점에서 도출하였다.

3.3 요인 간의 인과 분석

앞에서 도출한 영향요인들이 독립적으로 이슈에만 영향을 주는 것이 아니라, 요인들 간에도 영향을 줄 수 있다. A요인이 강화될 때 B요인도 같이 강화되면 긍정적인 영향(Positive Influence)을 A요인이 강화될 때 B요인이 약화되면 부정적인 영향(Negative Influence) 관계라고 정의한다.

이러한 요인간의 인과 분석에는 도메인 전문가 그룹 뿐만이 아니라 데이터 분석가와 시스템 다이내믹스 전문가도 참여하여 브레인스토밍과 피드백을 통해 진행되었고, 그 요인간의 인과관계 도식화를 [그림 3]에서 확인할 수 있다.

예를 들어, 특정 감염병의 전파력 요인이 강화되면 발병률이 높아지고, 병원체의 변이 가능성이 높아지며, 기존에 수립해놓은 예방법이나 치료법이 정상적으로 작동되지 않고 있으며, 발병국의 출국자 관리 통제 능력이 위험요소가 된다. 또한 이러한 전파력이 높은 감염병은 국내 입국자 관리 능력에도 부정적인 영향을 주며, 발병국과의 인적 교류 역시 위축될 수밖에 없다. 전파력이 높은 감염병의 발생은 해당 국가의 국민의 보건 인식의 수준을 높이며, 국제적인 언론의 관심이 높아져 실제 뉴스에 보도되는 건수도 증가하게 된다.

발병국과의 국제 교류 정도 요인이 증가하면 전파력, 발병률, 매개체의 이동(유입), 사람과의 접촉 가능성 요인들을 강화시키며, 국내 입국자 관리 통제 능력을 약화시킨다. 실제 2016년 8월의 브라질 리우 올림픽 사례를 생각해볼 수 있다. 올림픽으로 인해 브라질과의 국제 교류가 현저히 증가했고, 당시 여류인 브라질은 모기 매개 질환인 지카 바이러스의 전파력이 매우 높아져, 리우 올림픽의 일정을 변경하자는 논의까지 일어났다. 브라질 정부는 위생, 안전, 국민 보건 의식을 높이기 위해 정치, 경제적인 노력을 했으며, 당시의 국제 언론의 관심도 매우 높았다. 올림픽 출전 선수나 진행요원들이 숲모기에 물리지 않도록 주의를 당부했으며 국내에서도 브라질을 경유해서 국내에 입국하는 여행객들에 대해 집중적인 입국자 관리를 시행하였다. 또, 이 경우에는 매개체인 숲모기가 유입될 우려를 예상해 선박, 항공 등의 집중적인 방역 및 수입품에 대한 검역을 강화했다.

‘해외발생 감염병 국내 유입’ 이슈에 대한 영향요인 도출도 상당히 복잡하게 나왔지만 각 요인들 사이에 긍정적인 영향을 주는지, 부정적이 영향을 주는지를 분석해낸 인과지도 역시 직관적으로 파악하기 어려울 만큼 복잡하다는 것을 알 수 있다. 여기에서 가장 영향을 많이 주는 요인으로는 ‘병원체의 변이 가능성’, ‘발병국의 국가(정책)적 노력 및 효과’ 순으로 나타났다. 병원체의 변이는 그 가능성 정도도 예측이 어렵기 때문에 다른 요인에게 가장 많은 영향을 줄 수 있음을 알 수 있다. 다음으로는 감염병이 발병한 국가의 보건, 행정적인 정책의 수립 및 노력과 효과가 해당 감염병의 확산에 큰 영향을 주는 것임을 다시 한번 확인할 수 있다. 전세계의 교류가 점점 많아지고 있기에 국가 간의 보건 영역에서의 공조 역시 필요하다는 것을 확인할 수 있다.

가장 영향을 많이 받는 요인으로는 ‘발병률’, ‘전파율’으로 분석되었다. 이 두 요인을 효율적으로 통제, 조절할 수 있다면 다른 요인에게도 영향을 미친다는 것을 의미한다. 이와 같이 영향요인을 도출하고 요인간의 인과분석을 통해 인과지도를 그려보는 것은 이슈 탐지 기술 개발에 중점적으로 모니터링 해야 할 영향요인과 관련 변수들을 정의하는 계기가 될 수 있다.

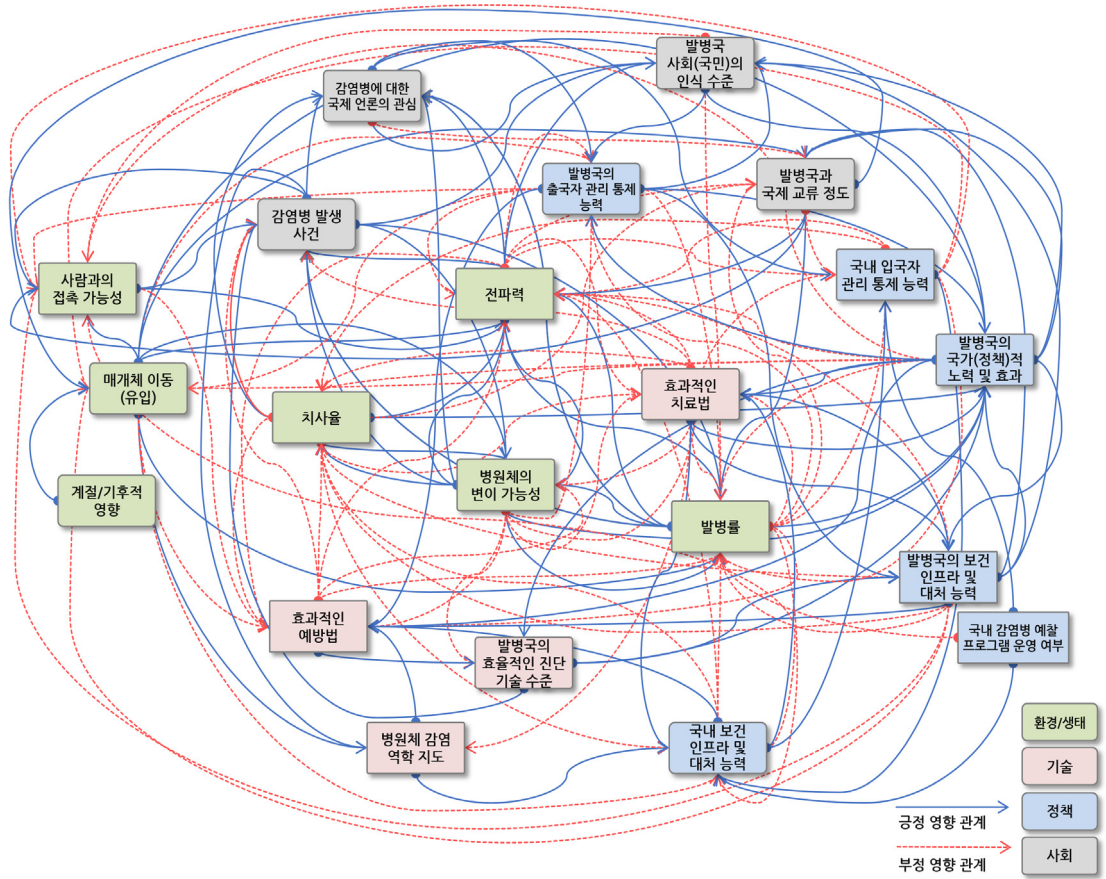


그림 3. ‘해외발생 감염병 국내 유입’ 이슈의 요인들의 인과관계를 분석한 인과지도

4. 결론

메르스, 지카 바이러스, 고병원성 조류 인플루엔자와 같은 국가, 사회의 안전을 위협하는 주요 이슈 해결은 신속한 정보의 수집, 분석과 이를 활용한 적절한 초기 대응이 무엇보다 중요하다. 하지만, 폭발적으로 증가하는 데이터는 이러한 활동의 핵심적인 문제를 바꾸어놓았다. 기존의 키워드 기반의 단편적인 분석만으로는 이슈의 탐지가 쉽지 않게 되었다.

본 논문에서는 이슈 탐지 기술의 기반 기술이 되는 도메인 지식 기반 이슈 탐지 모델을 설계하는 과정에 대해 기술하였다. ‘해외 발생 감염병의 국내 유입’이라는 이슈에 대해서 영향을 주는 요인들을 전문가 커뮤니티

를 통해 도출해내고, 이들 요인간의 인과분석을 통해 인과지도를 구성해내었다. 어떤 요인이 높은 영향 관계에 있는지, 어떤 요인을 중점 통제를 하면 타 요인에게 영향을 미칠 수 있는지를 확인할 수 있으며 이는 복잡한 영향요인 관계도를 정립했다는 것에 의미를 둘 수 있다.

이후에는 ‘해외발생 감염병의 국내 유입’ 인과지도를 구성하고 있는 변수에 실제 값을 부여하여 시뮬레이션을 해보고, 해외에서 발생한 감염병의 실제 사례 적용을 통해 모델의 검증은 진행할 필요가 있다. 이런 시뮬레이션을 통해 특정 감염병에 따라 가중치가 높아지는 요인과 변수를 선정하여, 본 지능형 이슈 탐지 모델을 이슈 탐지 기술에 적용할 예정이다.

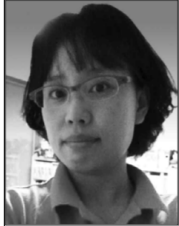
참 고 문 헌

- [1] 국립재난안전연구원, *복합적 미래예측방법론 분석을 통한 미래재난예측기법 개발*, 2013(12).
- [2] UKtransparency and CabinetofficeUK, *Open Data White Paper: Unleashing the Potential*, 2012.
- [3] 황미녕, 이승우, 최동진, “이슈 탐지(Horizon Scanning) 동향” *주간기술동향*, 2016.
- [4] B. G. Kim, S. Trimi, and J. Chung, “Big-Data Applications in the Government Sector,” *Commun. ACM*, Vol.57, No.3, pp.78-85, 2014.
- [5] 보건복지부, *2015 메르스 백서*, 2016.
- [6] 이근화, *효율적인 해외 감염병 모니터링 방안*, 질병관리본부, 2013.
- [7] C. Bahk, D. Scales, S. Mekaru, J. Brownstein, and C. Freifeld, “Comparing timeliness, content, and disease severity of formal and informal source outbreak reporting,” *BMC Infectious Diseases*, Vol.15, 2015.
- [8] G. Milinovich, R. Magalhaes, and W. Hu, “Role of big data in the early detection of Ebola and other emerging infectious disease,” *The Lancet Global Health*, Vol.3, 2015.
- [9] M. Keller, M. Blench, H. Tolentino, C. Freifeld, K. Mandl, A. Mawudeku, G. Eysenbush, and J. brownstein, “Use of Unstructured Event-Based Reports for Global Infectious Disease Surveillance,” *Emerging Infectious Diseases*, Vol.15, No.10, pp.689-695, 2009.
- [10] C. Bahk, D. Scales, S. Mekaru, J. Brownstein, and C. Freifeld, “Comparing Timeliness, Content, and Disease Severity of Formal and Informal Source Outbreak Reporting,” *BMC Infectious Diseases*, Vol.15, No.135, 2015.
- [11] W. Lipkin, “The changing face of pathogen discovery and surveillance,” *Nat Rev Microbiol*, Vol.11, No.2, 2013.
- [12] N. Collier, S. Doan, A. Kawazoe, R. Goodwin, M. Conway, Y. Taten, Q. Ngo, D. Dien, A. Kawtrakul, K. Takeuchi, M. Shigematsu, and K. Taniguchi, “BioCaster: detecting public health rumors with a Web-based text mining system,” *Bioinformatics*, Vol.15, No.24, 2008.
- [13] W. Sutherland and H. Woodroof, “The need for environmental horizon scanning,” *Trends Ecol. Evol.*, Vol.24, No.10, pp.523-527, 2009.
- [14] J. Ramos, “Anticipatory Governance : Traditions and Trajectories for Strategic Design,” *J. Futur. Stud.*, Vol.19, No.September, pp.35-52, 2014.
- [15] 박병원, “전략적 불확실성의 대응: 싱가포르 정부의 미래전략 수립 벤치마킹,” *Future Horizon*, Vol.25, pp.28-31, 2015.
- [16] Horizon Scanning Programme Team, <http://www.gov.uk/government/groups/horizon-scanning-programme-team>
- [17] B. Butland, S. Jebb, P. Kopelman, K. McPherson, S. Thomas, J. Mardell, and V. Parry, *Foresight. Tackling obesities: future choices. Project report*, Government Office for Science, 2007.
- [18] K. McPherson, T. Marsh, M. Brown, *Tackling obesities: future choices: Modelling future trends in obesity and the impact on health*, Department of Innovation, Universities and Skills, 2007.
- [19] Y. Low, K. Koh, S. Choo, C. Leng, L. James, L. Oh, and R. Wong, “Risk assessment and horizontal scanning scenario builder: developing a systematic approach to assess risk of importation of avian influenza(H5N1),” *Epidemiological News Bulletin*, Vol.35, No.1, pp.15-18, 2009.
- [20] M. Hwang, S. Lee, and H. Oh, “Building Issue-Data Maps to Support the Resolution of Socio-National Issues: Focusing on the Issue of increase in domestic inflow of foreign infectious diseases,” *Knowledge and Smart Technology*, pp.343-346, 2017.

저 자 소 개

황 미 녕(Mi-Nyeong Hwang)

정회원



- 2000년 : 부산대학교 전자계산학과(이학사)
- 2002년 : 부산대학교 전자계산학과(석사)
- 2002년 ~ 현재 : 한국과학기술정보연구원 융합기술연구본부

선임연구원

<관심분야> : 빅데이터, 시맨틱 웹, 데이터마이닝, 이슈 분석

이 승 우(Seungwoo Lee)

정회원



- 1999년 : 포스텍 컴퓨터공학과(석사)
- 2005년 : 포스텍 컴퓨터공학과(박사)
- 2006년 ~ 현재 : 한국과학기술정보연구원 융합기술연구본부

책임연구원

<관심분야> : 자연어처리, 정보추출, 시맨틱 웹, 정보 분석, 빅데이터