



외교안보 분야 빅데이터 분석

: 어떻게 수행할 것인가?

김상배 서울대학교 정치외교학부 교수
인디애나대학교 정치학 박사
전공분야: 정보혁명과 네트워크 세계정치
대표업적: 『버추얼 창과 그물망 방패: 사이버 안보의 세계정치와 한국』(근간), 『아라크네의 국제정치학: 네트워크 세계정치이론의 도전』(2014), 『정보혁명과 권력변환: 네트워크 정치학의 시각』(2010), 『정보화 시대의 표준경쟁: 윈텔리즘과 일본의 컴퓨터산업』(2007) 외 다수

〈 초 록 〉

최근 빅데이터는 큰 규모를 활용해서 작은 규모에서는 불가능했던 새로운 통찰과 새로운 형태의 가치를 창출해내는 일을 가능케 하고 있으며, 다양한 분야에서 적용되면서 그 유용성을 인정받고 있다. 그러나 빅데이터 분석을 외교안보정책에 적용하려는 움직임은 여타 분야에 비해서 매우 미흡한 실정이다. 이 글은 빅데이터를 활용한 외교안보정책의 효과적 추진을 위해서 필요한 전제조건과 실현방안을 검토하고, 실제로 이를 추진할 경우 원용되어야 할 빅데이터 분석의 프레임워크를 제시하였다. 이 글은 크게 두 부분으로 구성되었다. 제2장은 빅데이터를 활용한 외교안보정책의 의미를 살펴보았으며, 아울러 여타 분야와는 달리 외교안보 분야의 빅데이터 분석이 지니는 특성을 살펴보고, 이를 토대로 실제로 빅데이터를 활용한 외교안보 정책을 수행하기 위해서 필요한 분석 프레임워크를 제시하였다. 제3장은 외교안보 분야 빅데이터 분석 프레임워크를 7단계 전략모델로 제시하고 각 단계에서 고려해야 할 변수들을 살펴보았다. 각 단계는 i) 쟁점 확인, ii) 질문 세팅과 대상 타겟팅, iii) 수집 및 처리, iv) 분석 및 피드백, v) 시각화 활용, vi) 결과 해석, vii) 정책결정 등으로 나누어진다. 특히 각 단계별로 외교안보 분야의 빅데이터 분석이기 때문에 특별히 고려해야 할 문제들을 지적하고 이들 문제를 해결하는 구체적인 방안들을 검토하였다. 맺음말에서는 빅데이터를 활용한 외교안보 정책결정 시스템 구축을 위해서 필요한 향후 조치에 대해서 살펴보았다.

1. 머리말

디지털 경제와 정보사회의 확산에 따라 규모를 가늠할 수 없을 정도로 많은 데이터가 생산되면서 출현한 빅데이터 환경은 각 분야에 큰 변화를 야기하고 있다(최버거·쿠리어, 2013; 정용찬, 2013; 매일경제 기획팀·서울대빅데이터센터, 2014). 빅데이터는 큰 규모를 활용해서 작은 규모에서는 불가능했던 새로운 통찰과 새로운 형태의 가치를 창출해내는 일을 가능케 하고 있으며 최근 다양한 분야에서 적용되면서 그 유용성을 인정받고 있다. 비즈니스 부문에서 빅데이터의 활용은 이미 널리 확산되었으며(채승병, 2015), 공공행정, 선거전략, 재난관리 부문에서도 빅데이터를 활용하는 사례가 늘어나고 있다. 예를 들어, 전자정부, 복지행정, 의료행정(신종플루, 구제역, 에볼라), 교육행정, 조세행정, 경찰행정, 교통행정, 선거 캠페인 등의 분야에 공공 데이터, 특히 오픈이언 마이닝, 다양한 웹문서 분석, 댓글 분석 등이 활용되고 있다. 이들 분야에서 신속하고 급박하게 이루어져야 하는 실시간 의사결정을 보조하는 시스템의 일부로서 빅데이터 분석의 결과를 활용하는 유용성이 인정되고 있다. 경험적 직관에 의존하던 기존의 아날로그 의사결정 시스템을 넘어서 과학적 예측과 패턴에 대한 분석을 바탕으로 한 디지털 의사결정 시스템으로 이행되는 양상이 나타나고 있다.

오늘날 외교안보 문제는 점점 더 복잡해지고 그 불확실성이 커지고 있으며 대응시간도 점차 짧아지고 있다. 이러한 맥락에서 아날로그적 경험에 기초한 ‘직관적 대응 외교’를 넘어서 디지털 시스템을 활용한 ‘과학적 예측 외교’의 유용성에 대한 관심이 커지고 있다. 외교정책의 결정 및 실행의 효율과 속도를 향상시키기 위해서는 외교안보 빅데이터 시스템을 구축하고 활용하는 ‘빅데이터 외교’에 대한 기대가 커지고 있다. 기존의 방식으로 읽히지 않는 숨은 의도나 의미구조를 읽어내는 데 있어 빅데이터 분석을 활용하는 것은 일정한 의미가 있을 것으로 판단된다. 과거 스몰 데이터(small data) 환경에서는 외교관의 경험적 지식이나 개인적 네트워크를 활용한 정보의 수집과 분석을 바탕으로 외교안보 정책이 입안 및 실행되었다. 빅데이터 환경에서는 텍스트 기반의 데이터뿐만 아니라 웹사이트나 블로그, 소셜 미디어 등을 통해서 생성되는 큰 규모의 정형 및 비정형 데이터를 처리하여 새로운 통찰에 바탕으로 둔 의사결정의 기초자료를 도출하는 것이 가능해 졌다. 요컨대, 빅데이터 환경의 출현으로 인해서 기존의 HUMINT(human intelligence) 기반의 외교안보 정책결정으로부터 OSINT(open source intelligence)를 기반으로 한 새로운 정책결정으로서의 패러다임 변화가 발생하고 있다.

그럼에도 빅데이터 분석을 외교안보 정책에 적용하려는 움직임은 아직까지 여타 분야에 비해서 미흡한 실정이다. 최근 빅데이터가 국제정치 분야에서 지니는 의미와 그 활용 가능성에 주목하는 연구들이 나오기 시작하고 있다(Zwitter, 2015; Chandler, 2015; 강선주, 2014; 김상배, 2015; 배영자, 2015; 이승주, 2015; 박종희, 2015). 또한 최근 안보연구의



시각에서 빅데이터와 첩보, 감시의 문제를 다루거나, 또는 빅데이터 시대의 국가 주권의 변화에 대한 문제제기를 하는 학술연구들이 나오기 시작하고 있다(Roderick, 2014; Esposti, 2014; Skelly and Eichenmüller, 2015; Hansen and Porter, 2015; 조현석 편, 2013). 외교안보 및 통일국방 분야에 빅데이터 분석을 활용하는 사례들도 속속 등장하고 있다. 예를 들어, 미국 CSIS(Center for Strategic and International Studies)에서 북한의 핵실험과 미사일 발사 등 WMD 활동에 대한 빅데이터 분석을 행하고 있다. 국내 정부부처 차원에서도 통일부의 북한 정세 분석 시스템의 구축 및 가동, 국방부 TICN(첩보 분석 시스템), 다출처 정보 융합처리, SoSA 복합체계 분석 등을 수행하고 있다. 그러나 본격적으로 외교안보 정책을 보조하는 시스템으로서 빅데이터 분석을 활용하는 방법에 대한 연구는 거의 없는 실정이다.

이러한 맥락에서 볼 때 빅데이터를 외교안보 정책결정과정에 활용하는 작업의 의의와 실현 가능성을 검토하고, 여타 공공정책 분야에서 빅데이터를 활용한 사례들을 외교안보 분야로 원용하는 작업을 펼치는 의미가 있다. 현재 학계의 연구현황을 고려할 때 본격적으로 빅데이터를 활용하여 실제로 외교안보 정책에 적용 가능한 정책 모듈(policy module)의 개발에 앞서 이러한 시도의 가능성과 한계 및 고려사항을 살펴보는 기초연구가 필요하다. 이에 이 글은 빅데이터를 활용한 외교안보 정책의 효과적 추진을 위해서 필요한 전제조건과 실현 방안을 검토하고, 실제로 이를 추진할 경우 원용되어야 할 빅데이터 분석의 프레임워크를 제시하였다. 이와 같이 외교안보정책에서 빅데이터의 효용성을 제고하기 위한 분석 프레임워크 및 매뉴얼 개발의 가능성을 진단하는 연구는 여러 가지 면에서 그 효과가 기대된다. 특정 이슈와 관련하여 해당 국가에서 형성·진화하고 있는 대(對) 한반도 정책 담론 환경을 파악하는 정책자료로 활용 가능하며, 더 나아가 특정 이슈와 관련된 해당국 정부의 정책에 대한 효과적인 대응방안을 도출하는 시나리오 개발도 가능할 것이다.

이 글은 크게 두 부분으로 구성되었다. 제2장은 빅데이터를 활용한 외교안보 정책의 의미를 살펴보았으며, 아울러 여타 분야와는 달리 외교안보 분야의 빅데이터 분석이 지니는 특성을 살펴보고, 이를 토대로 실제로 빅데이터를 활용한 외교안보 정책을 수행하기 위해서 필요한 분석 프레임워크를 제시하였다. 제3장은 외교안보 분야 빅데이터 분석 프레임워크를 7단계 전략모델로 제시하고 각 단계에서 고려해야 할 변수들을 살펴보았다. 각 단계는 i) 쟁점 확인, ii) 질문 세팅과 대상 타깃팅, iii) 수집 및 처리, iv) 분석 및 피드백, v) 시각화 활용, vi) 결과 해석, vii) 정책결정 등으로 나누어진다. 특히 각 단계별로 외교안보 분야의 빅데이터 분석이기 때문에 특별히 고려해야 할 문제들을 지적하고 이들 문제를 해결하는 구체적인 방안들을 검토하였다. 끝으로 맺음말에서는 빅데이터를 활용한 외교안보 정책결정 시스템 구축을 위해서 필요한 향후 조치에 대해서 살펴보았다.



II. 빅데이터 분석과 외교안보 정책

1. 외교안보 분야 빅데이터 분석의 의미

외교안보 분야 빅데이터 활용은 그리 활성화 되어 있지는 않지만, 최근 외교안보 문제의 복잡화에 따라 빅데이터를 활용하는 수요가 증가하고 있다. 빅데이터 분석 및 활용을 통해 국가간 외교 네트워크를 종합적·입체적으로 시각화해서 선제적으로 파악하는 것이 가능하다. 또한 해외 민간대중을 대상으로 하는 정부 및 비정부 행위자들의 공공외교 활동에서 빅데이터의 활용은 더욱 중요해지고 있다. 방대한 양의 데이터를 다루기 때문에 보다 많은 변수들을 복합적으로 고려할 수 있고 시간 범위도 중장기 관점으로 확장해서 볼 수도 있다. 빅데이터의 공공외교적 활용은 외교자원의 합리적 분배 및 최적화에도 긍정적 영향을 미쳐 외교의 생산성을 향상시킬 수 있을 것으로 기대된다. 빅데이터 활용의 유용성 중 하나는 즉시성에 있으며, 이는 빅데이터 분석이 일반 설문조사에 비해서는 훨씬 더 빠르게 여론동향을 읽어내고 정책적인 대응을 가능하게 한다는 기대를 낳고 있다. 이러한 맥락에서 기존의 외교부 조직 내의 데이터 활용 문제 등을 모두 고려하여 현재 거론되는 빅데이터 외교의 위상을 설정하면 <그림-1>과 같은 논의를 펼칠 수 있다.

<그림-1> 외교안보 분야의 디지털 데이터 활용

	조직 내부	조직 외부
정형 데이터	<1-영역> 외교 정보화	<3-영역> 지식외교
비정형 데이터	<2-영역> 외교 암묵지	<4-영역> 빅데이터 외교

출처: 송영조(2015)에서 응용; 김상배(2015), p.23

첫째, 조직 내부의 정형 데이터를 외교 분야에 활용하는 디지털 외교이다(<그림-1>의

<1-영역>). 인터넷이 널리 확산되지 않았던 웹1.0시대에도 외교행정이나 여권업무 등과 같은 조직 내부 업무를 위해서 정형 데이터를 디지털화해서 활용하는 ‘외교 정보화’ 사업이 진행되었다(김상배, 2004). 이렇게 인트라넷을 활용하여 내부에서 작성되고 수집된 자료를 체계적으로 관리하고 공유함으로써 외교업무를 효율화하는 능력은 앞으로도 중요할 것이다. 외교부의 경우에는 해외공관에서 대면접촉을 해서 전문을 쓰는 경우가 많은데, 완전히 대외 비로 취급되는 텍스트 자료들을 프로그램을 통해 데이터로 축적하여 분석하는 것이 가능할 것이다.

둘째, 조직 내부의 비정형 데이터를 외교 분야에 활용하는 디지털 외교이다(<그림-1>의 <2-영역>). 이는 외교 암묵지(tacit knowledge)를 기록하고 디지털화해서 활용하는 능력과 관련된다. 이미지 자료, 간단한 메모나 팩스서신 등의 텍스트 자료, 녹취록과 같은 구술 자료, 외교관들의 경험, 노하우, 현지 정보, 국제기구에 대한 정보, 그리고 타국 외교관에 대한 개인접촉 정보 등은 비정형화된 형태로 존재할 수밖에 없다. 여태까지는 수집과 축적 자체가 어려웠던 이러한 비정형 데이터들을 체계적으로 관리, 저장, 공유하는 작업이 데이터 기술의 발달에 의해서 주목받고 있다.

셋째, 조직 외부의 정형 데이터를 외교 분야에 활용하는 디지털 외교이다(<그림-1>의 <3-영역>). 조직 외부의 정형 데이터란, 다른 국가의 정부, 민간기업, 국제기구, 비정부기구 등에서 생성하는 공공 및 민간 데이터를 말한다. 이러한 데이터들은 정보혁명의 진전과 함께 그 양과 범위가 폭발적으로 증가할 것으로 예측된다. 이런 맥락에서 다양한 데이터에 대한 접근과 수집, 분석을 위한 체계적인 플랫폼을 마련하여 실제 정책연구와 개발에 활용하는 ‘지식외교’가 주목받고 있다. 특히 디지털 정보를 활용하는 지식외교는 당초 아날로그 정보에 의존하는 전문가들의 직관에 주로 의존했던 외교정책 결정자들의 업무에 큰 변화를 줄 것이다.

끝으로, 조직 외부의 비정형 데이터를 외교 분야에 활용하는 디지털 외교이다(<그림-1>의 <4-영역>). 인터넷과 소셜 미디어의 확산은 외교 분야에서도 상상할 수 없는 규모의 비정형 데이터를 생산하고 있다. 외교부 내에 개설한 웹사이트를 통해서 양방향 대민서비스를 벌이거나 다양한 디지털 기기를 통해 수집된 비정형 빅데이터의 패턴을 읽어서 외교정책에 활용할 필요성이 커지고 있다. 이 글에서 강조하는 ‘빅데이터 외교’는, 앞서 언급한 세 가지 유형과 모두 관련되지만, 주로 이렇게 조직 외부에서 생성되는 다양한 비정형 데이터를 활용하는 유형을 염두에 두었다.

빅데이터 분석의 매력은 외교부 조직 내의 데이터를 처리하는 것보다는 오픈 플랫폼으로 바깥에 널리 있는 데이터를 분석하는 데 있다. 데이터나 정보 하나하나가 아니라 전체 구도를 봄으로써 그 전에 보이지 않는 패턴을 찾을 수 있다는 것이 빅데이터 분석의 매력이다. 따라서 조직 내의 텍스트를 데이터 분석하는 것과 조직 바깥의 오픈 데이터를 분석하는 두

가지 작업을 동시에 수행해야 할 것이다. 조직 내부의 데이터는 경험이 많은 외교관이 다룰 수도 있겠지만, 빅데이터 분석의 도움을 받으면 시간을 단축해서 입체적으로 효율성이 있는 데이터를 가공해낼 수 있다. 그러나 외부 데이터, 오픈 플랫폼에서 얻는 데이터들을 통해 외교관들이 쉽게 떠올릴 수 없는 이머징 이슈나 예외사항(outlier)에 대한 것 또는 통상적인 직관으로는 잡히지는 않지만 그동안 보지 못했던 것을 보게 해주는 효과가 있을 것이다.

2. 분석 프레임워크의 개발

빅데이터를 활용한 외교안보 정책결정 시스템의 구축을 위해서는 기술공학적 프레임워크를 넘어서는 국제정치학적 프레임워크의 개발 및 이를 실제로 수행하는 과정에서 참조할 일종의 매뉴얼 개발이 필요하다. 기술적인 자산이나 툴이 그 자체로 충분히 효율을 주는 분야가 있고, 그렇지 않은 분야가 있는데, 외교안보 분야는 후자에 속한다. 따라서 기술적 도구가 즉각적인 효율을 주지 못하기 때문에, 여러 가지 맞춤화(customizing)가 필요하고, 어떤 도구를 사용할 것인지를 문제 이전에 분석 프레임워크의 개발이 중요하다. 다시 말해 외교안보 분야에서는 “우리 브랜드에 대해서 소비자들이 어떻게 생각하고, 경쟁자들의 전략이 어떤 트렌드를 보이는가?” 와 같이 경영학 분야에서 사용하는 단순한 질문이 아니라 상당히 복잡한 질문을 던질 수밖에 없고, 실제로 역동적으로 움직이는 현상에 대해서 포착할 필요가 있다.

외교안보 분야 빅데이터 활용의 분석 프레임워크와 관련하여 참조할 수 있는 것으로는 미래부에서 만든 빅데이터 표준화 플랫폼(Procedure of Big Data) 6단계(kbig.kr 참조)가 있는데, 이는 아래와 같다.

- i) 분석의 목적을 정의: 무엇을 얻을 것이냐, 어디에 사용할 것인가?
- ii) 분석데이터가 무엇인가? 내부 데이터와 외부 데이터의 구분, 구조적(structured) 데이터와 비구조적(unstructured) 데이터의 구분
- iii) 분석방법: 어떤 분석 방법으로 할 것인가?
- iv) 실질적인 분석: 수집-처리-분석-활용: 이 부분은 기술공학적 분석의 핵심을 이룸
- v) 분석결과에 대한 해석: 직관의 발견
- vi) 분석결과의 활용: 정책결정을 지원

이러한 6단계에 대한 논의 중에서 네 번째 단계에서 거론한 데이터의 수집-처리-분석-활용을 4단계로 나누어 정리하면 <그림-2>와 같다.

〈그림-2〉 수집-처리-분석-활용의 4단계



출처: 이상원(2017), p.9

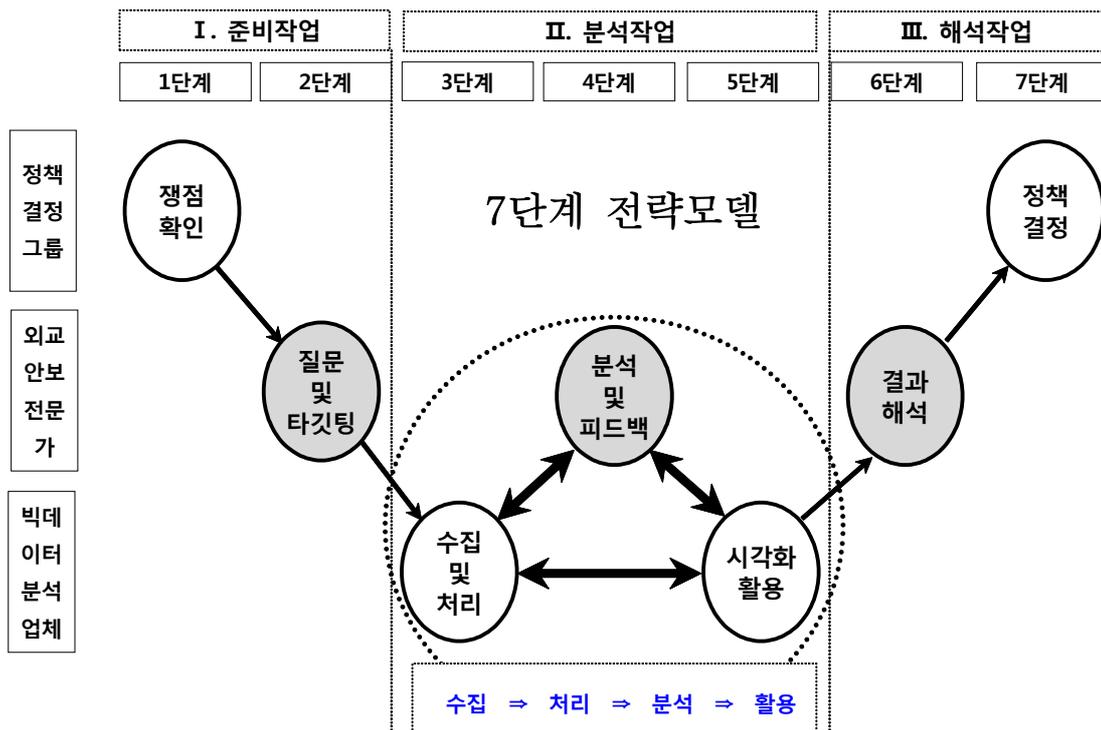
그러나 기술공학적 차원에서 빅데이터를 활용하는 시도가 외교안보 분야에서 효과를 보기 위해서는 단순히 빅데이터 시스템 개발 등 기술공학적 접근에만 그쳐서는 안 되고, 외교안보 시각을 원용한 빅데이터의 분석 및 활용이 필요하다. 다시 말해 기술적·공학적 시스템 설계의 문제를 넘어서 이를 가이드하는 국제정치학의 직관적 개입과 방향 설정 및 수집된 빅데이터에 대한 해석의 작업이 필요하다. 기술적으로 구현할 수 있는 부분들을 넘어서기 위해서는 특정 단계가 아니라 전 과정에 걸쳐서 데이터에 대한 재검토와 피드백 작업이 상당부분 요구된다. 방대한 양의 데이터 소스를 찾고 시스템을 구축해, 그 안에서 자연어를 처리하고 분석하는 과정에서 끊임없이 국제정치학적 직관이 투입되어야 한다. 예를 들어, 기술공학적 과정을 통해서 도출된 결과를 실제로 유용하게 활용하기 위해서는 상대국의 외교정책과 관련된 빅데이터 분석의 대상(인물, 키워드, 웹사이트, SNS, 기관)을 명확히 하는 것이 필요하다. 또한 우리의 직관에 반하는 정보에 대한 해석이 필요하며, 이러한 관점에서 볼 때, 국제정치학자들이 어떠한 해석적인 개입을 하는가가 중요하다. 예를 들어, 여러 가지 데이터 가운데 어떤 것을 걸러낼 것인지, 오류정보(misinformation)와 거짓 뉴스(fake news), 의도적으로 제공한 정보 등을 걸러내는 과정이 중요하다. 이러한 문제의식을 바탕으로 빅데이터의 활용과 연계되는 국제정치학적 분석모델에 대한 논의가 필요하다.

이러한 점에서 볼 때 외교안보 분야에서 수행하는 빅데이터 분석에 있어서는 아래와 같은 네 가지 접근을 원용하여야 한다. 첫째, 유력인사 기반 접근(influencer-based approach)인데, 이는 일반 대중이 아닌, 외교 이슈에 영향을 미칠 수 있는 인물 및 기관의 담론 분석을 통한 커뮤니케이션 전략을 반영한다. 둘째, 이슈 중심의 접근(issue-specific

approach)인데, 이는 단일한 기술 플랫폼 중심이 아닌, 각 외교 이슈에 적합한 개별화된 분석과 모니터링을 위한 DB 플랫폼을 구축한다. 셋째, 전략적 문제 해결 접근(strategic problem-solving approach)인데, 이는 통상적 동향보고가 아닌, 외교 전략 입안을 위한 질문에 기반을 두어 해답을 찾고자 하는 분석 프레임워크를 축적한다. 끝으로, 구성주의적 커뮤니케이션(constructivist communication)인데, 이는 현실주의에 대한 맹목적 추종이 아닌, 전략적 틈새를 발굴하고 적극적으로 커뮤니케이션하기 위한 방안을 모색한다(김도훈, 2017).

이러한 외교안보 분야의 특성을 고려하여, 빅데이터를 활용할 경우 필요할 것으로 예상되는 작업들의 형식과 내용의 로드맵을 담은 분석 프레임워크(또는 작업 매뉴얼)의 7단계 전략모델을 제시하면 <그림-3>과 같다. 다시 말해 i) 쟁점 확인 → ii) 질문 및 타겟팅 → iii) 수집 및 처리 → iv) 분석 및 피드백 → v) 시각화 활용 → vi) 결과 해석 → vii) 정책결정 등의 7단계를 따라서 외교안보 분야에서 빅데이터를 정책공공외교에 활용하는 방안을 제안한다. 이러한 7단계를 좀 더 크게 구분하면 준비-분석-해석 작업의 3단계로 나눌 수 있다. 이 중에서 3-5단계는 데이터 공학 분야에서 원용하는 수집 → 처리 → 분석 → 활용의 4단계를 품고 있다(이상원, 2017). 외교안보적 통찰을 강조하는 이 글에서는 이러한 4단계에 더하여 세 단계를 더 추가하였다.

<그림-3> 빅데이터 활용의 분석 프레임워크





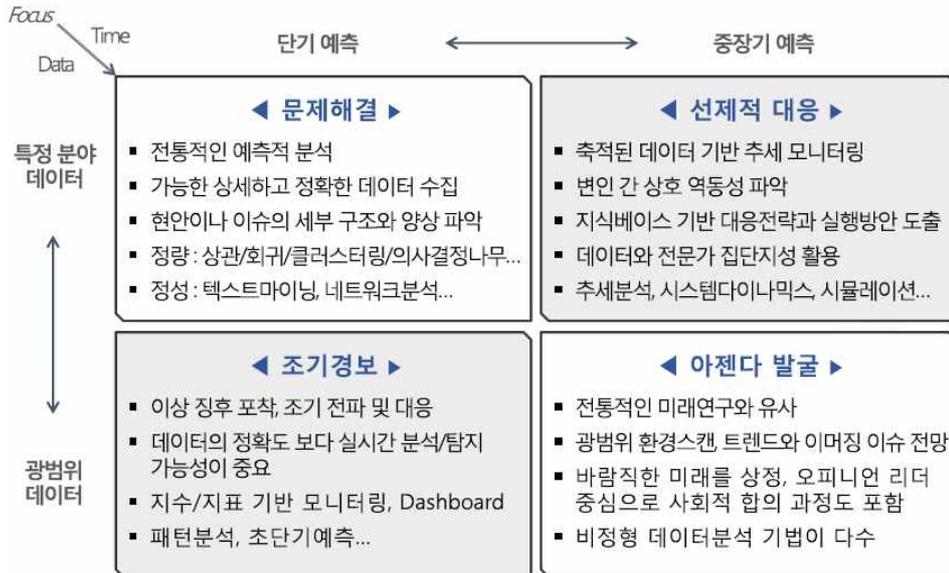
이러한 7단계 전략모델은 정책결정그룹과 외교안보전문가 및 빅데이터 분석 업체의 3자간의 협업 네트워크를 바탕으로 하여 작동한다. 이 과정에서 정책결정그룹은 i) 쟁점확인과 vii) 정책결정을 담당한다. 외교안보전문가 그룹은 ii) 질문 및 타깃팅, iv) 분석 및 피드백, vi) 결과해석의 단계를 담당하며 기술공학적 분석에 국제정치학적 직관을 가미하는 핵심적 역할을 담당한다. 빅데이터 분석업체는 iii) 분석 및 처리와 v) 시각화 활용의 단계를 담당하며 이는 데이터 공학에서 말하는 수집 → 처리 → 분석 → 활용의 네 단계에 해당하는 작업이다. 이러한 7단계 전략모델에 입각한 분석 프레임워크는 외교안보 분야에서 빅데이터를 활용한 정책결정의 보조 시스템을 마련하기 위한 편이를 위해서 세분화한 것이며, 좀 더 크게 구분하면 외교정책 결정그룹과 외교안보 전문가들이 담당하는 준비작업과 해석작업 단계, 그리고 빅데이터 분석 업체가 담당하는 분석작업의 셋으로 나누어 볼 수도 있다.

Ⅲ. 외교안보 빅데이터 분석의 7단계 전략모델

1. 쟁점 확인

이 단계는 정책결정 그룹의 정책수요에 따라 쟁점을 확인(issue identification)하는 단계이다. 쟁점 확인의 주기는 정기적(월간, 격주간 또는 주간)으로 하는 트랙과 사안별로 그때그때 제기하는 두 가지 방식을 병행할 필요가 있다. 이러한 과정에서 가장 중요한 것은 어떤 이슈를 빅데이터 분석의 대상으로 해야 할 것인지를 결정하는 문제이다. 예를 들어, 중요하면서도 반복 가능성이 있는 이슈들(예를 들어 북핵, 사드 등 안보 이슈, 한류 등 소프트 파워 이슈, 경제통상 이슈 등)의 트렌드를 확인하는 방안이 있다. 갑자기 발생한 사건의 성격과 그에 잇달아 발생할 가능성이 있는 사건 및 관련 당사자들의 대응양식 등을 예측하는 방안도 있다. 또는 잠재하고 있지만 조만간 부상할 것으로 예상되는 창발적 이슈(emerging issues)를 확인하는 방안도 있다. 그러나 쟁점 확인의 방법은 고정된 것이 아니라 계속해서 개작(adaptation)이 이루어져야 하고, 전체적으로 분석하려는 타깃 대상이나 범위도 유연하게 조정하는 것이 필요하다.

〈그림-4〉 빅데이터 분석의 목적 설정



출처: 이영주(2017), p.21

이러한 쟁점 확인의 과정에서는 수행하려는 빅데이터 분석의 목적을 설정하는 작업이 병행되어야 한다. 시간 축을 따라서 단기예측과 중장기 예측, 그리고 데이터의 성격에 따라서 특정 분야 데이터와 광범위 데이터로 나눌 수 있다. 이를 바탕으로 빅데이터 분석의 목적을 설정하면, 〈그림-4〉에서 제시한 바와 같이, 문제해결, 선제대응, 조기경보, 어젠다 발굴 등의 넷으로 나누어 볼 수 있다. 확인되는 쟁점의 성격에 따라서 빅데이터의 분석 목적을 명확히 하는 것이 중요하다.

이러한 문제의식을 바탕으로 향후 외교안보 분야에 빅데이터를 활용 가능한 쟁점으로 서 미국의 사례와 관련된 것만 뽑아 보면 다음을 예로 들 수 있다. 외교안보 분야의 경우 트럼프 행정부가 북한의 핵실험과 미사일 발사 등의 도발을 억지하기 위해서 취할 수 있는 정책옵션과 방법에 대한 빅데이터 분석을 수행할 수 있다. 이밖에도 한국 정부의 남북정상회담 추진, 북한의 대륙간탄도 미사일 발사 실험, 미국의 한반도 전술핵무기 배치, 미국의 한반도 사드 배치 철회 결정, 사드 관련 중국 내 협한감정 추이, 한반도 통일에 관한 중국 내 여론 등을 사례로 들 수 있다. 경제통상 분야의 경우 트럼프 행정부가 한미 FTA 협상을 재추진할 가능성과 재협상 이슈 등에 대한 빅데이터 분석을 수행할 수 있다. 이밖에도 한미 무역수지 불균형, 한미 FTA 재협상, 한중 FTA, 한국의 환율조작국 지정 여부 등을 들 수 있다. 공공 외교 분야의 경우, 미국의 정책결정자나 미 국민들의 한국의 국가브랜드에 대한 이미지, 예를 들어 한국의 새 정부에 대한 호감도 빅데이터 분석 등을 수행할 수 있다. 워싱턴 DC 서클

에 한국에 대한 이미지, 한국 교육시스템에 대한 이미지 등도 좋은 빅데이터 분석의 사례들이다.

2. 질문 및 타깃팅

외교안보 분야는 매우 유동적이고 광범위하기 때문에 구체적인 질문과 타깃팅이 필요하다. 제2단계의 작업은 전체 빅데이터 분석 작업에서 가장 중요한 의미를 지니는 단계로서, 분석을 원하는 해당 하위분야 전문가들과의 협업을 바탕으로 진행되어야 한다. 외교안보 정책에서 빅데이터 활용전략은 막연하게 관련 데이터를 수집하고 분석하는 것이 아니라 명확히 분석대상을 타깃팅(심지어 마이크로 타깃팅까지도 포함)해서 구체화·세분화하는 것이 필요하다. 다시 말해, 해당 하위 분야의 이슈들마다 타깃이 되는 사람과 조직, 데이터 플랫폼이 모두 다를 수밖에 없다. 즉 인물, 기관 및 매체 등 특정한 외교목표와 대상, 이슈에 따른 타깃 세분화에 기반을 둔 외교안보 ‘정책모듈(policy module)’을 개발하고 빅데이터 활용과 연계시켜야 한다. 이들 분석대상은 기본 그룹을 정해 놓고 질문과 사안에 따라서 조정하여야 한다. 이러한 맥락에서 제2단계는 크게 i) 질문 세팅, ii) 키워드 선정, iii) 인물 타깃팅, iv) 미디어 타깃팅 등의 네 가지 하위 단계로 나누어 볼 수 있다.

1) 질문 세팅

통상적 동향보고가 아닌, 외교안보 전략의 현안분석, 즉 전략적 문제 해결 접근(strategic problem-solving approach)이 목적인만큼 분석질문의 제시가 매우 중요하다. 빅데이터 분석에 있어서는 ‘답을 찾는 것’보다 ‘질문을 만드는 것’이 중요하다는 지적이 제기된다. ‘어떠한 종류의 질문을 제시할 수 있는가?’의 문제인데, ‘똑똑한 질문’을 던져야 ‘똑똑한 대답’이 나온다. 비유컨대, 빅데이터 분석은 현문현답(賢問賢答)의 영역이다. 질문을 세팅하는 방법과 관련하여 초기 단계에는 큰 주제로 시작해서 점차적으로 세분화하는 것이 필요하다. 처음에는 ‘이것과 어느 대상이 상관관계가 있는가?’와 같은 일반적인 질문을 던지거나, 직관적 관찰로는 상관관계의 실마리가 보이는 데, 지금은 데이터가 충분치 않아 본격적인 분석은 할 수 없지만, 향후 데이터가 축적되면 분석할 수 있을 것으로 예상되는 주제를 확인하는 질문을 던지거나, 아니면 초기 단계이지만 상호 간의 상관관계를 확인할 수 있는 것은 확인해 보는 종류의 질문을 던질 수 있다. 대략 주제확인형 질문 → 상관관계 확인형 질문 → 인과관계 확인형 질문 등의 3단계를 거쳐서 질문을 던질 수 있을 것이다.

또한 질문을 세팅할 때 이미 아는 문제를 확인하는 질문(즉 트렌드 분석) 이외에도 창발형 이슈 검출용 질문을 던지는 방법(즉 이미지 이슈 분석)을 생각해 볼 수 있다. 예를 들어

2017년 한반도 주변정세의 트렌드를 묻는 질문을 던지고 그 과정에서 나타나는 패턴을 빅데이터 분석을 통해서 확인하는 연구가 있을 수 있다면, 이렇듯 한반도 주변정세와 관련하여 이미 알고 있는 질문이외에도 이른바 ‘창발적 이슈(emerging issue)’에 대한 질문을 병행해서 세팅하는 것도 필요하다. 창발적 이슈란, ‘현재 직접적인 영향을 끼치는 것처럼 보이지만 장차 새로운 트렌드(trend)로 부상할 수 있는 이슈’를 의미한다. 여태까지 이머징 이슈의 탄생에 대한 예측은 직관적이고 선형적이었는데, 이를 데이터 기법을 활용하여 객관화시켜 보자는 시도이다. 아직 두드러지게 부상하지는 않았지만 직간접적으로 계속 회자되고 있는 이머징 이슈가 있다면 이를 미리 파악하고 선제적으로 대응할 수 있을 것이다.

2) 키워드 선정

질문이 세팅되고 난 다음에는 크롤링 작업에 사용할 키워드를 선정하는 작업이 필요하다. 이러한 키워드 선정 작업은 각 외교안보 이슈에 적합한 개별화된 분석과 모니터링을 목적으로 하는 이슈 중심의 접근(issue-specific approach)이기 때문에 더욱 중요하다. 키워드 선정의 방식은 임의적이고 자의적으로 선정하는 것은 아니고, 문헌정보학에서 원용하고 있는 방법론을 적용하여야 정확한 분석결과를 도출할 수 있다. 예를 들어 북한 핵실험 및 미사일 발사 등과 관련된 미국의 정책을 분석하기 위해서는 대략 10-20여개의 키워드를 선정할 수 있는데, 이 과정에서 동일한 이슈라도 키워드를 어떻게 선정하느냐가 중요하다. 북한의 미사일을 로켓으로 부르기도 하고, 인공위성이라고도 하는데, 어느 하나가 선정되는 순간 다른 20개의 연관이 되는 키워드 사이에서 상이한 결과값이 도출될 것으로 예상된다. 또한 ‘실험’이라고 부르는 경우와 ‘도발’이라고 부르는 경우도 차이가 있다. 이에 대해서는 어떤 용어를 선택할 것인지를 미리 정해 놓을 수는 없고 각 연구수행 과정에서 연구수행자가 상황에 따라서 판단해야 한다.

단어가 너무 일반적인 단어이거나 혹은 분석에 방해되는 단어일 경우 이를 전처리 과정에서 불용어 리스트를 작성하여 제외할 수 있다. 사실 동일한 내용의 단어이기 때문에 제외하고 보아야겠다는 것은 결국은 연구자의 판단에 의해서 결정되는데, 연구목표에 따라서 특정 단어를 제외하고 보면 전체 그림이 달라질 수 있다. 따라서 해당 이슈 전문가의 식견이 필요한 대목이다. 형태소 분석 시에 불용어에 해당되는 단어를 내보내는 작업을 실시한다. 여러 가지의 버전을 두고, 어떤 것이 더 객관적인지를 비교하는 작업도 병행해야 한다. 과제를 진행할 때 키워드 선정, 전처리, 노이즈 제거 등에 보통 시간과 비용이 상대적으로 많이 들어간다. 이러한 과정이 사실상 전체 과제에서 약 80% 정도를 차지한다고 보아야 할 것이다.

하나의 키워드 트렌드에 대해서 찾는 것은 쉬울 수 있으나, 그 자체만으로는 유의미한 결과값을 얻어내기 어려울 수 있으므로 키워드와 관련된 토픽을 선정하고, 관련 토픽의 세부

항목을 살펴보는 방법을 취할 필요가 있다. 이러한 맥락에서 키워드를 선정하기 전에 먼저 주제별 카테고리를 정하는 것이 필요한데, 카테고리별로 중요한 핵심단어들을 선정하는 것이 유용하다. 예를 들어, 북한과 관련해서는 경제, 정치, 생활 등의 카테고리를 설정할 수 있고, 정치 분야에서는 국방, 외교 등 카테고리를 잡을 수 있으며, 그 다음 단계로는 카테고리별 주요 단어를 설정하는 작업의 순으로 진행할 수 있다. 한편 키워드 선정에 있어서 명사가 아닌 형용사를 선정하고 이를 구별하는 시도를 벌여야 할 필요가 있을 수도 있다. 예를 들어, 오피니언 마이닝을 수행할 경우 감성분석도 하는데, 주로 명사를 가지고 분석을 한 것에서 더 나아가서 감정이 실리는 형용사 분석까지 하고 있다. 좀 더 구체적으로는 감정이 실리는 형용사가 어떤 명사를 꾸미는지의 문제까지 분석하는 단계에까지 데이터 분석기법이 진전되어 있음에 주목할 필요가 있다.

실제 빅데이터 분석을 수행할 경우 키워드들 간의 관계를 설정하는 인덱싱(indexing) 작업이 필요할 수도 있다. 분석대상으로 삼고 있는 것에 대해서 키워드와 관련된 유사 단어의 가치 및 그들 간의 위계관계까지도 설정해야만 이에 맞춰서 분석할 수 있다. 키워드를 단 순나열만 해놓으면 출현빈도 분석까지밖에 안 되며, 범위를 좁히고 내용을 구체화시켜 줘야만 제대로 분석을 할 수 있다. 초기 작업 단계에서부터 엄청난 데이터를 다 넣어서 분석하기 보다는 전략적인 시나리오와 그에 따라서 카테고리별 주요 단어들을 선정해 주고, 이미 데이터가 있는 역사적 사건과 비교하는 분석을 해보는 것도 방법이다. 또한 샘플 의존성(sample dependancy)의 문제를 고려해서 동일한 샘플을 너무 오래 사용해도 부적절하므로 여러 가지 모형으로 분석을 시도할 필요가 있다.

3) 인물 타깃팅

외교안보 빅데이터 분석은 일반 대중이 아닌, 외교안보 이슈에 영향을 미칠 수 있는 인물 및 기관의 담론을 분석하는 유력인사 기반 접근(influencer-based approach)을 취하고 있기 때문에 분석대상이 되는 인물이나 타깃 그룹에 대한 리스트 업이 필요하다. 일단은 어떤 사람들이 중요한지를 밝혀내는 것이 필요하며, 오피니언 리더들의 네트워크 지도를 만들고, 해당 이슈에 대한 찬반그룹(pro-con)을 구분해서 어떤 사람이 핵심인지를 밝히는 작업이 요구된다. 이를 토대로 상이한 사안별로 파워 엘리트인 개인과 기관의 리스트 업 및 DB 구축이 필요하다. 외교안보 분야에서 파워 엘리트가 누구냐를 밝히는 문제는 사전적으로 타깃팅을 할 수도 있지만, 분석을 통해서 사후적으로 발견하는 방식을 따를 수도 있다. 그러나 실제로는 오프라인에서 파워 엘리트들이 누구인지를 사전적으로 리스트 업하고, 이들을 기초자료로 해서 리스트를 구체화시켜 나가는 것이 적중률을 높이는 방법이 될 수 있다. 이러한 기초자료를 기반으로 네트워크 분석을 하는 도중에 생각지 못했던 유력인사를 사후적으로 발견하는 효과를 노리는 방법을 병행하는 것이 유용하다.

연구대상 국가 및 이슈에 대해서 각각 인물 데이터베이스를 마련하는 것이 필요한데, 이는 해당 이슈 전문가들의 경험과 식견을 바탕으로 해야 하는 중요한 문제이다. 대중의 트렌드를 파악하는 것과 특정인이나 특정기관의 성향을 파악하는 것은 완전히 다를 것 같은데, 특정인을 대상으로 하는 작업은 어느 정도 전문가의 직관에 의존하는 선택의 과정이 필요하며, 따라서 그 선택이 얼마나 정확한지에 대한 합리적 의심을 할 수밖에 없다. 빅데이터 분석은 특정기관과 특정인이 맺고 있는 관계망으로부터 유추되는, 스스로는 드러내고 싶지 않지만 간접적으로 표출되어지는 부분을 포착하겠다는 것이 목적이기 때문에 애초에 전문가의 직관을 바탕으로 그 특정인을 제대로 선정하는 것이 중요하다. 인물 리스트 업을 할 때 발생할 수 있는 오류를 해결하기 위해서 작은 규모의 기초자료에서 출발해서 몇 달 간의 상호작용을 지켜보고 지속적으로 표본집단을 늘려가는 눈덩이 표집(snowball sampling) 방식으로 보완할 수 있다. 역으로 더 많은 인물 리스트를 마련하고 작업을 진행하면서 줄여나가는 방법도 가능하다.

인물 타깃팅의 구체적인 방법을 살펴보면, 미국 정계에서 한반도 정책에 영향력이 있는 파워 엘리트들 100여명 표본집단으로 뽑아 보는 작업을 수행할 수 있다. 표본집단으로 선정할 정책서클에서 만들어지는 여론이 궁극적으로는 백악관이나 미국 대통령에 영향을 미칠 것일 것이므로, 이들 정책서클의 여론을 빅데이터 분석하는 의미가 있다. 예를 들어, 대통령과 백악관(관료) 참모진 10여 명, 한반도 외교현안에 관련된 의회 인사 10-20여 명, 한반도 관련 연구기관(브루킹스, 랜드, 외교협회 등) 외교전문가 20-30여 명, 뉴스미디어의 한반도 전문기자 20-30여 명, 미국의 정치 블로거 10여 명 등으로 구성된 대략 100여명의 표본집단에서 분석을 시작할 수 있다. 더 나아가 이들을 성향(매파/비둘기파)에 따라서 분류하여 분석하거나, 유명 로비스트나 로비기관, 언론기관 등을 추가로 데이터 소스로 지정하는 세분화 및 확장 방법을 취할 수 있다. 이를 기반으로 각 인물들의 주장, 이익 등을 분석·비교하여 무엇이 의도하는 분석에 적절한지를 판단한다. 온라인 설문을 보내거나 워크숍을 하거나, 인터뷰나 토론 등을 해보고 그 결과에 기반을 두어 그들과의 네트워크를 계속 확대해 나가는 구성주의적 커뮤니케이션(constructivist communication) 접근을 가미하면, 분석의 효과를 높일 수 있다.

일반 대중이 아니라 전문가들의 견해에 초점을 맞추는 작업이기는 하지만, 사안에 따라서는 일반 대중의 여론을 완전히 무시할 수 없는 측면도 있을 것이다. 따라서 지나치게 의존하지 않는 범위 내에서 일반 대중의 여론도 참고용으로 검토할 필요가 있다. 일반대중과 전문가를 대상으로 할 때 데이터를 가져오는 방식은 동일하지만, 분석하는 데 있어서는 차이가 있을 수 있다. 일반 여론의 경우에는 컴퓨터가 처리한 결과의 핵심을 파악하는 방식으로 진행하면 되고, 전문가들의 의견은 상대적으로 적은 양이기 때문에 프로세싱 결과 값에 더해서 직접 연구진이 검토하여 해석하는 과정이 필요하다.

이미 스포 데이터 환경 하에서도 언론 모니터링과 주요 인물 추적은 하고 있는데, 외교안보정책의 시각에서 빅데이터를 활용해서 추가적인 분석을 하는 것이 어떤 부가가치를 창출할 수 있을지를 밝혀야 한다. 기존 방식과의 차별화 지점은, 유력인사들은 자신의 견해를 쉽게 바꾸지 않기 때문에 직접적인 접촉만으로 그 의도 변화를 파악하거나 또는 더 나아가 설득하기는 어렵지만, 빅데이터 분석은 그 이면에 있는 내용을 파악할 수 있다는 점이다. 빅데이터 분석을 활용하여 해당 유력인사와 친분관계를 맺고 있어 간접적으로 해당 유력인사의 의도를 파악할 수 있거나, 더 나아가서는 해당 유력인사에 영향을 줄 수 있는 주변 사람들을 찾아내서 그 관계적 맥락을 분석하는 방법이 가능하다. 공적인물(public figure)에게는 기본적으로 특정한 직함이 있는데, 뉴스 매체에서 발신자 검색이 된다는 점을 활용할 수 있다. 예를 들어, ‘누구의 말에 의하면’ 같은 표현 등을 통해서 SNS를 전혀 사용하지 않는 인사들의 견해를 포착하는 데 활용할 수 있다. 이러한 기법을 활용하여 특정 외교정책 이슈에 대한 선제적 파악과 예방적 대응 등의 조치가 가능한지를 검토할 수 있을 것으로 기대된다.

4) 미디어 타깃팅

이 단계는 타깃팅한 인물과 그룹의 활동에 대한 데이터와 정보가 담겨 있는 미디어를 선정하는 단계이다. 정부, 의회, 미디어, 싱크탱크 등의 사이트와 게재된 i) 정부기관의 공식 발표, ii) 각 기관의 정형 텍스트 자료, iii) 비정형 자료 등이 대상이 된다. 특히 미디어의 경우에는 좀 더 세분화된 타깃팅이 필요한데 기성 뉴스 미디어 이외에도 페이스북이나 트위터와 같은 소셜 미디어, 블로그 등에 따라서 DB구축이 필요하다. 정부기관 사이트에 담긴 공식적인 문건, 발언(statement)과 소셜 미디어 데이터를 함께 보는 것이 유용하다. 정책적인 목적을 달성하기 위해서, 상대방의 반응을 탐지할 수 있는 데이터로는 언론보다는 정책서클의 SNS에 더 많이 주목하는 것이 효과적일 것으로 예상된다. 앞서 언급한 표본집단에 속한 사람들이 온라인에서 SNS를 이용해서 연결되어 있는 상태를 분석하고, 추가로 이들이 자주 참고하는 여타 미디어의 샘플 수를 늘려나간다. 분석의 효과성을 높이기 위해서는 이상의 분석 대상을 교차 및 중첩하여 데이터를 수집할 필요가 있다.

최근 소셜 미디어만 분석하면 신뢰를 얻기 어렵기 때문에 기타 미디어를 복합적으로 교차해서 보아야 한다는 인식이 대두되고 있다. 소셜 데이터뿐만 아니라 정량화되어 있는 다양한 지표들을 활용하는 방법을 복합하는 것이 중요하다. 이들 정량지표와 매칭해서 소셜 미디어를 대상으로 한 빅데이터 분석을 사용할 수 있는 방법을 고려해 볼 필요가 있다. 소셜 데이터에만 의존하는 것이 아니라 복합적인 데이터가 필요하다는 점은, 신문 매체에서도 일반 기사와 오피니언을 구분해서 보는 것이 필요하다는 인식과도 맥이 닿는다. 태생적으로 중립적 객관성을 지향하는 매체(즉 신문의 일반기사)와 주관적 주장과 인식의 자율성과 이들의 상호 경합 속에서 오피니언을 파악하려는 매체(소셜 미디어) 간에 발생하는 차이를 정확히

이해하는 것이 중요하다. 결국 이러한 변수들을 모두 고려하여 전체 데이터 세트를 어떻게 정하고 어떻게 수집할 것인가를 결정할 수밖에 없다.

이러한 점들을 염두에 두고 트럼프 행정부의 대한반도 정책을 엿보기 위해서 타깃 인물과 타깃 미디어를 연결해서 설정할 경우 다음과 같은 다섯 가지 그룹을 고려해 볼 수 있을 것이다.

- i) 주요 관료들의 SNS 계정, 개인 블로그 유튜브, 페이스북, 트위터 등
- ii) 미국 주요 기관들의 웹사이트나 페이스북 계정 등
- iii) 이러한 정계 소식에 초점을 맞추어서 뉴스를 다루거나 정보를 유통시키는 주요 매체(예를 들어 Politico 등), 외교 전문매체 등
- iv) 한미관계나 국제관계를 다루는 전문 연구소나 연구공동체, 싱크탱크, 미국 전문가들의 블로그 등
- v) 민간 차원(한 개인일수도 있고 시민단체나 NGO일 수도 있음)의 블로그, 일종의 정치외교 마니아들(일종의 파워 블로거)의 블로그나 SNS 등

빅데이터 분석의 대상을 확장해 나갈 경우 이슈별, 국가별로 차별화 및 범위조정도 필요하다. 다시 말해, 국가마다 미국처럼 정책결정자들이 SNS를 하는 경우도 있지만, 전혀 그렇지 않고 정책결정자들이 누구인지 기초적인 정보조차 얻기 어려운 국가들도 있다. 예를 들어, 미국무장관인 힐러슨은 페이스북을 쓰지만, 중국의 외교부장인 왕이는 페이스북을 안 쓴다. 특히 중국의 경우 개인이 자유롭게 표현할 수 없는 환경이기 때문에 미국의 경우에 비해서 공식매체들이나 기관 계정의 데이터를 더욱 더 많이 사용하게 될 가능성이 크다. 또한 중국의 경우에는 선전성 매체의 경우는 공개를 하지만, 이것만으로 유의미한 분석을 할 수 없다. 선전성 매체 이외의 경우에는 분석을 통해 의미를 이끌어낼 수 있는 데이터들도 많겠지만, 그러한 데이터를 실질적으로 수집하지 못할 수도 있다. 이상의 어려움에도 불구하고 미디어의 보급과 사용 및 접근성 등의 문제는 국가별로 나타나는 차이로 인정한 상태에서 빅데이터 분석을 수행할 수밖에 없다.

3. 수집 및 처리

외교안보 정책 이슈에 기반을 둔 각각의 타깃팅이 완료되면, 축적된 DB와 분석프레임의 요구에 부합하는 빅데이터 분석 프로그램의 개발 및 알고리즘 설계 및 개작이 필요하다. 외교안보 전문가들이 빅데이터 분석에 필요한 일종의 규칙(rule)과 가설을 제시하면, 이를 처리할 수 있는 소프트웨어나 기술을 활용하여 크롤러를 짜고, 이를 활용하여 필요한 데이터

를 수집하여 자연어 처리를 수행한다. 현재 상황을 고려할 때, 빅데이터 분석의 기술 시스템 개발은 i) 대형업체로부터 솔루션을 구입 사용, ii) 중소기업체와 연구팀의 협업, iii) 연구팀의 자체 개발 등의 방법이 있다. 아직 외교안보 분야의 빅데이터를 정책 적용의 목적으로 가지고 활용하는 방법에 대한 선행연구가 거의 없으므로 시행착오의 과정을 거쳐서라도 한국의 실정에 맞는 빅데이터 분석의 방법론을 개발하는 의미가 있다. 이러한 점을 염두에 두고 향후 진행될 외교안보 분야 빅데이터 분석은 i) 데이터 수집, ii) 데이터 처리, iii) 데이터 저장의 세 단계를 거쳐서 진행하여야 할 것이다.

1) 데이터 수집

실시간 크롤러인 웹 크롤링, SNS 크롤링, META 서치 등과 같이 이미 개발된 프로그램을 활용하여 데이터의 수집 작업, 즉 크롤링(crawling)을 수행한다. 데이터 수집은 빅데이터 분석에서 제일 핵심적인 작업이라고 할 수 있으며, 데이터를 수집할 수 없으면 아무 것도 할 수 없기 때문에, 어떤 데이터를 어떻게 모을 것인가의 문제는 가장 근본적인 부분이다. 아무리 좋은 기법이라도 데이터를 잘못 수집하거나 왜곡된 데이터가 포함되면 전혀 다른 결과가 나올 수 있으므로 데이터 수집이 중요하다. 국내외에 이미 수집이 되어 있거나 처리까지 되어 있는 데이터 또는 SNS나 미디어 등을 긁어서 얻은 데이터 등을 같이 결합해서 활용하는 방법들을 사용할 수 있다. 예를 들면 국내에서는 통계청의 K-데이터 중에서 외교 관련 데이터, 언론사의 웹 데이터들이 있고, 이들 데이터를 기반으로 외교안보 전문가들의 직관을 투영하여 분석 주제를 설정할 수 있다.

데이터 소스별로 구별하여 인터넷 사용 패턴이 국가별로 다른 점에 착안한 데이터 수집이 필요하다. 최근 지역별로 소재(location)한 데이터를 수집하는 것이 어려워지고 있는데, 이는 최근 데이터 자체의 가치가 점차 높아지고 있어 데이터를 무상으로 제공해주지 않는 추세 때문이다. 예를 들어, 한 때 데이터를 개방했던 트위터도 요즘에는 외부유출을 막아놓은 상태이기 때문에, 데이터를 어떻게 확보하는지가 문제가 된다. 서버에서 크롤러가 데이터를 수집하고 있다는 것이 감지되면 트래픽에 제한이 걸리거나, 보안상 이슈로 인해서 데이터를 수집하지 못할 수도 있다. 업체들은 ‘긁히는 것만 긁어 와서 작업을 해도 괜찮다’ 라고 할 수도 있지만, 이러한 이유 때문에 데이터가 편향될 가능성이 있음을 유념해야 한다.

또한 이 단계에서는 기술적 문제뿐만 아니라 분석대상 기관이나 국가의 개인정보 및 지적재산권 관련 국내법 또는 국제법 등의 저촉 여부에 유의할 필요가 있다. 다시 말해, 해외 소재의 빅데이터 활용 시 개인정보 보호, 데이터 소유권 및 주권 침해 문제 등과 관련하여 해당국의 국내법과 국제법에 대한 법적·규범적 검토가 필요하다. 외교안보 분야 빅데이터 분석은 대체로 공개 데이터, 즉 널리 있는 데이터들을 수집해서 새로운 가치를 만들겠다는 것이기 때문에 대체로 법적으로는 문제가 되지 않는다. 그러나 결과물을 어떻게 사용하는지에

따라서 법적인 문제도 얼마든지 생길 수 있음에 유의해야 한다. 처음에 요청한 용도가 아니라 다른 용도로 사용하는 경우에는 법적 문제가 발생할 수 있으며, 공개적으로 이러한 작업을 하게 됨에 따라서 야기될 수 있는 문제를 항상 염두에 두어야 한다. 또한 공개되어 있거나 식별 자료가 없는 데이터라고 하더라도 지시해 놓은 권리사항을 위반하거나 활용과정에서 수집대상 주체의 의사에 반한다는 일이 발생한다면 얼마든지 법적인 문제가 제기될 수 있다. 따라서 공공 데이터 사용 시에도 조심스러운 부분이 있으며, 경우에 따라서는 알고리즘을 투명하게 공개해야 할 필요성이 발생할 수도 있다.

미디어별 차이를 이해하는 것도 중요한데, 트위터의 경우는 완전히 오픈된 소셜 미디어이기 때문에 수집된 데이터는 공공 데이터가 되므로 법적인 문제가 없다고 주장하는 것이 가능하며, 그것이 일반적인 관행이다. 그렇지만 페이스북 개인계정의 경우는 도덕적·윤리적·법적 이슈가 잠재적으로 발생할 수 있다. 기술적으로는 가능할지 몰라도 데이터 수집 시에 발생하는 법적 문제 때문에 대부분의 업체들이 원칙적으로 사용하지 않고 있고, 만약 개인계정이 아니라 오픈된 페이지가 있다면 이 경우의 데이터는 수집할 수 있다. 만약 개인계정을 사전에 동의 없이 수집해서 활용한 것이 대외적으로 출판되는 결과물에 반영된다면 법적인 문제가 발생할 수 있다. 그러나 아직 현실적으로 공개된 데이터를 활용해서 법적인 분쟁으로 비화된 사례는 없다. 객관적으로 오픈된 데이터만 수집하는 경우, 그 중에서 데이터의 문제가 있다면 데이터를 제공하는 쪽에서의 문제일 가능성이 크기 때문에 악의적이고 불법적으로 사용하지 않는다면 특별히 법적인 문제는 없을 것이라는 것이 법학자들의 견해이다.

2) 데이터 처리

인간이 직관적으로 읽어내지 못하는 부분에 대해서 인공지능 프로그램을 활용하여 패턴을 읽어내는 빅데이터 분석을 수행한다. 빅데이터 분석에서 백미는 굉장히 많은 데이터에서 특정한 패턴을 읽어내는 것, 즉 사람이 물리적으로 읽을 수 있는 한계를 넘어서 소프트웨어를 활용하여 데이터를 처리하는 것이다. 최근 데이터 처리에는 하둡 기반 비정형 처리, 정형 처리(DW), 실시간 소셜 데이터 처리(mongoDB-키워드 추출, 특성추출 등) 등과 같은 다양한 방식이 이용되고 있다. 데이터 처리를 위해서는 자연어 처리, 국제어 처리, 높은 수준의 검색기술, 데이터 커버 등이 요구되는데, 현재 시장에 나와 있는 업체들은 대부분 이를 제공할 수 있는 기술력을 갖추고 있는 것으로 판단된다. 그러나 이러한 지표는 실제로 외교안보 분야 빅데이터 분석의 필요조건에 불과하며, 관건은 외교 분야에서의 분석에 있어서는 기본적인 기술적 분석을 넘어서 수집된 데이터를 목적에 맞게 구조화하고 해석하는 능력이다. 또한 이를 가능하게 하는 분석 프레임워크를 구축하고 그 진행과정에서 컨설팅 인사이트(consulting insight)를 제공하는 것이 중요하다. 협업과 커뮤니케이션에서 가장 중요한 것

은 외교안보 분야의 특수성과 메커니즘을 이해할 수 있는 국제정치학적인 직관이다.

외교안보 영역의 경우, 전 세계의 어떤 데이터를 타깃팅할 것인지에 따라서 언어 문제가 있을 수 있고, 언어에 따라서 인덱싱(indexing)이 달라질 수 있는데, 실질적으로 이를 완전히 해결할 수 있는 업체는 없는 실정이다. 그러나 어려움은 있을 수 있지만 큰 문제는 아니라는 것이 업계 관계자들의 견해이다. 한국어와 일본어가 유사하기 때문에 상대적으로 쉽게 접근하는 것이 가능하고, 중국어는 언어 자체보다는 중국이 폐쇄적인 편이기 때문에 데이터를 어떻게 모을 것인지 정도가 문제 될 수는 있다. 실제로 삼성전자의 경우, 한국어, 영어, 일본어로 5천 만 개의 기술문서를 분석하고 경쟁자들과의 비교도 하는 시스템을 구축하였다. 국내 업체는 한국어, 영어, 일본어, 중국어 네 개의 언어를 분석 가능한데, 현재 한반도 주변 4국 언어 가운데 러시아어 분석이 다소 미흡한 것으로 알려져 있다. 그러나 모든 플랫폼은 언어 의존 부분과 언어 비의존 부분이 있기 때문에, 러시아어도 언어인식 프로그램만 있으면 그 이후의 과정은 거의 동일하게 진행할 수 있다는 것이 업계의 중론이다.

3) 데이터 저장

빅데이터 분석에서는 수집·처리한 데이터를 DB나 아카이브 등에 저장하는 단계도 매우 중요하다. 우선은 미리 아카이브를 구축하는 것이 필요한데, 아직까지는 누구도 외교안보 분야의 빅데이터에 대한 아카이브를 마련하고 있지 않은 것으로 판단된다. 만약 북핵 6차 핵실험까지 관련된 데이터의 아카이브를 구비하고 있는 경우, 7차 핵실험을 예측하는 것에 대한 분석의 비용은 훨씬 줄어들 것이다. 몇 개라도 선행 사례가 있으면 분석의 난이도가 훨씬 낮아지게 되기 때문이다. 업체들이 구글 검색을 사용해도 됨에도 불구하고 굳이 따로 크롤러를 두는 것은 기존의 데이터들을 지속적으로 추적·저장하는 것이 필요하기 때문이다. 또한 최근 데이터 저장에 드는 비용이 크게 줄어들음으로 인해서 빅데이터를 저장할 필요성은 더욱 커지고 있다.

데이터를 저장하고 DB를 구축하는 과정은 중장기적으로 정보를 축적해서 언제든지 분석할 수 있도록 준비하는 프레임워크를 갖추는 의미가 있다. 실시간 정보 공유와 커뮤니케이션이 중요한 SNS는, 당장의 글로벌 이슈나 한국 관련 이슈가 급부상하였을 때 미국 정치권에서의 반응을 실시간 분석할 수 있는 대상이지만, 이러한 SNS에 대한 분석을 더욱 강력하게 만드는 것은 축적된 DB의 존재 여부이다. 처음 DB를 구축하는 작업에 시간이 많이 걸리겠지만, 일단 축적이 되면 그 다음부터는 상대적으로 비용과 시간이 적게 소요될 것이다. 일종의 학습과정 또는 진화를 거치면 자동화까지 가능한 프로그램이 만들어질 수 있다. 따라서 시뮬레이션하는 단계와 축적하는 단계, 자동화하는 단계 등으로 점진적으로 나아가갈 필요가 있다.

4. 분석 및 피드백

데이터의 분석은 데이터 공학의 시각에서 본 기술공학적 분석과 국제정치학의 시각에서 본 외교안보적 분석의 두 가지 측면으로 나누어진다. 기술공학적 차원에서는 수집·처리·저장된 데이터를 분석하기 위해 다양한 방법들이 동원되는데, 텍스트 마이닝의 방법으로서 다중분류, 군집분석, 감성분석, 이슈/트렌드 분석, 연관분석 등이 있다. 이 중에서 가장 쉽게 생각해 볼 수 있는 분석방법은 키워드의 출현빈도 분석이다. 주로 이머징 이슈에서 가장 간단하게 적용할 수 있는 것이 언급의 횟수로 분석하는 것인데, 즉 사람들이 얼마나 그 검색어를 많이 사용하는가를 살펴보는 것이다. ‘First Story Detection’ 분석은 주로 신문 기사에 적용이 되는데, 헤드라인 뉴스가 얼마나 새로운 이슈에 대해서 언급하고 있는지를 추적·탐지해 나가는 기법이다. 언급 횟수에 따라서 중요성을 파악하는 방법의 장점은 가장 손쉽게 사용할 수 있다는 것이지만, 단점은 과연 언급을 많이 했다고 해서 반드시 중요한가, 언급을 하는 것이 과연 대중들의 순수한 의도이나, 아니면 다른 의도에 의해서 언급 횟수가 올라가게끔 하는 다른 기제가 있었느냐를 알기 어렵다는 데 있다.

따라서 단순히 언급 횟수만으로 분석하는 것은 문제가 있을 수 있다. 특히 단어만을 분석하기 때문에 문맥을 파악할 수가 없다는 한계가 있다. 단어는 핵심어구로 표현될 때 의미가 있는 것이지, 단어 자체의 횟수만으로는 한계가 굉장히 많다. 그러나 이는 언론매체를 일일이 스크랩하는 것이 아니라 디지털화해서 요약된 정보를 획득할 수 있다는 점에서 유용하다. 단기간 내(예를 들어 1일)의 분석도 가능한데, 그것이 가능하려면 과거의 데이터를 통해서 학습을 하고 난 다음에는 가능하다. 하루에 나오는 데이터 양이 한정되어 있기 때문에, 하루 단위의 분석은 몇 시간 안에도 충분히 가능하다. 일종의 DB가 구축되어 있으면 추가로 들어오는 데이터는 많지 않기 때문에 그 절차를 자동화시키면 하루 단위의 분석도 가능하다.

이러한 점에서 문맥(context)을 고려한 분석방법을 동원할 필요가 있다. 예를 들어, 특정 대상에 대해서 언론이 긍정적으로 언급하는지, 부정적으로 언급하는지를 실시간으로 모니터링하고, 어떤 키워드들이 핵심적으로 등장하고 그 비중이 어떻게 되는지, 이것이 실시간으로 어떻게 변화하는지를 살펴볼 필요가 있다. 문맥을 함께 보는 분석의 대표적인 예로, 특정 단어가 어떤 단어와 함께 출현하는지를 보는 동시출현횟수 분석이 가장 많이 사용된다. 그러나 이 방법도 문제점이 없지 않다. 예를 들어, 1,000자의 텍스트로 된 블로그 게시글이 있는데, 첫 번째 단어와 마지막 단어가 동시에 출현했다고 해서 횟수를 1로 할 때, 단순히 동시출현했다는 것이 문맥적으로 연관된 의미를 내포하고 있는지는 쉽게 단정할 수 없기 때문이다. 한 윈도우(한 문장을 의미) 안에서의 어휘 변화를 보는 방법도 원용할 수 있다. 텍스트 마이닝에서 문장은 인간이 자신의 사고나 주장을 담아낼 수 있는 기본적인 단위인데, 한 문

장이나 앞뒤 문장까지 문맥이 들어가는 문맥정보가 있어야 한다. 문맥정보 없이 동시출현횟수만 가지고 분석을 할 때는 한계가 있을 수밖에 없다. 이밖에도 단어를 주제별로 군집화해서 분석하는 모델로서 토픽 모델링도 있다.

이러한 기술공학적 분석이외에도 이러한 분석 과정을 외교안보적 시각에서 분석하여 데이터를 ‘구조화’ 하려는 노력이 가미되어야 한다. 선행연구와 구조화, 협업의 문제는 연동되어 있기 때문에 데이터를 기술적으로 분석하는 것이 아니라 정치외교 전문가들과 기술자들이 협업해서 무엇을 구조화할 것인지를 원칙을 찾아내는 작업이 필요하다. 미국의 외교안보 정책을 분석하는 경우 그 기저에 깔린 국내정치 문제 등을 어떻게 반영해서 데이터를 추출할 것인가의 문제 등은 해당 분야 전문가의 도움을 받지 않을 수 없다. 빅데이터라고 해서 데이터가 많을수록 좋다고 생각할 수도 있지만, 정치외교 분야에 맞는 원리나 원칙을 찾아내기 위해서는 오히려 스톱 데이터에 집중해야 할 수도 있다. 이러한 단계를 밟아나가는 데 있어서 통상적인 동향 분석이 아니라 구체적인 질문을 던지고 그에 대한 해답을 찾아나가는 정책현안 분석이 필요하다.

이상의 기술공학적 및 외교안보적 분석의 과정을 통해서 도출된 분석 결과를 피드백의 과정을 통해서 지속적으로 검토하는 작업이 필요하다. 필요에 따라 3-4단계를 반복 수행하여 프로그램의 오류와 편견(bias)을 수정하는 작업을 수행하여야 한다. 일종의 분석 매뉴얼이나 프레임워크를 구축한 다음에는 시뮬레이션과 예측작업도 해 보고, 가설에 의해서 정확도가 높으면 채택하는 것이고, 정확도가 떨어지면 가설을 보완하는 방향으로 조정해 나가야 한다.

5. 시각화 활용

분석된 결과를 프리젠테이션하는 과정에서는 이른바 인포그래픽스라고도 불리는 다양한 시각화 기술을 활용할 수 있다. 현재 많이 원용되는 시각화 기술로는 D3 기반 시각화(area차트, tree차트, line차트, circle차트), BI 기반 시각화, 실시간 소셜 시각화(토픽 트렌드, 토픽 클라우드, 토픽탱크, 문서량 트렌드), 개발자 포털(API 연계, APP 툴 테스트) 등이 원용되고 있다. 이러한 툴을 잘 다루는데도 시간이 많이 걸리며, 보통은 장면(shot) 개념으로 데이터 분석을 하는데, 상호작용이 가능한(interactive) 인포그래픽스를 제공하는 것이 중요하다. 이러한 시각화의 유용성은 기술공학적 분석 결과를 외교안보 전문가와 정책결정자들이 쉽게 이해할 수 있는 형식으로 시각화 작업을 수행한다는 데 있다.

6. 결과 해석

이 단계는 빅데이터 분석을 수행한 후, 해당 이슈 외교안보 분야의 국제정치 전문가들이 분석결과를 해석하는 단계이다. 빅데이터 분석을 통해서 지금까지 인지하지 못하고 있었던 사실을 얼마나 확인할 수 있는지의 여부가 가장 중요할 것이다. 새로운 아이디어를 주는 것이 아니라 지금까지 알고 있던 사실에 대한 백(back) 데이터에 불과하게 된다면 과연 의미 있는 분석이라고 할 수 있는지를 물어야 할 것이다. 그렇지만 마찬가지로 지금까지의 지식과는 다른 결과가 나올 경우, 이를 어떻게 설득할 것인가도 또 다른 차원의 문제를 야기하는 것이 사실이다. 결국 분석 결과에 대한 제대로 된 해석을 위해서는 외교안보 전문가들의 ‘상황지성(contextual intelligence)’ 이 중요할 것이다.

궁극적으로 분석결과에 영향을 주는 정보와 그 메커니즘을 이해하는 것이 중요하다. 통계를 내는 것은 누구나 할 수 있으나, 그 통계를 통해서 어떠한 실질적인 의미를 찾아낼 것 인지가 핵심이다. 빅데이터 분석이 기대치 않은 결과(unexpected outcome)를 얻는 것이 목표라고 하는데, 만약 정책결정자 입장에서 자신의 직관이나 경험지와 전혀 다른 결과가 나왔을 때 단순히 정책결정자가 판단해야 할 문제로 남기는 것이 아니라 일종의 가이드라인을 줄 수 있어야 할 것이다. 결국 빅데이터 분석 결과를 그대로 따라야 한다는 것이 아니라, 이런 결과가 나왔고 이런 측면이 있을 수도 있다는 점을 참고하라는 의미로 분석결과를 제공하여야 할 것이다. 다시 말해 만약 경험지나 직관과 다른 결과가 나왔을 경우라도 보완적으로 활용할 수 있는 ‘상황지성’ 을 보유하고 있어야 할 것이다. 따라서 앞서 제시한 이슈들에 대한 분석을 수행함에 있어서, 예측하기 전에 성찰(reflection)하는 과정을 밟는 것이 중요하다. 특히 예측하고자 하는 경우에는 시스템 리소스를 활용해서 회고적인 분석을 잘 하는 것이 필요하다.

7. 정책결정

분석결과 활용은 절대적인 의미가 있는 것이 아니라 의사결정의 보조시스템으로 활용하는 의미가 있음에 유의할 필요가 있다. 외교안보 분야에서 빅데이터를 활용하여 얻을 수 있는 효과는 절대적이라기보다는 외교정책 결정자의 직관적 통찰력을 높일 수 있는 ‘참조모델’ 로서 가치가 있다. 다시 말해, 외교안보 분야의 정책결정 영역에서 데이터 분석을 ‘결정모델’ 로 활용한다는 것은 다소 무리가 있으며, 제한된 범위 내에서 전략적 함의를 도출하는 정도로 활용하는 의미는 있다. ‘경험지’ 에 ‘분석지’ 를 더하는 형태의 전략모델을 개발한다는 의미로 접근해야 할 것이다. 정책결정 보조자료의 작성은 두 가지 트랙(two-track) 이상으로 진행하는 것이 좋다. 반복적으로 일어나는 이슈와 급작스럽게 일어나

는 사건이나 X-이벤트를 구분해서 진행하는 방법을 고려할 수 있다. 보조자료의 제공은 월례 보고서, 주례 리포트, 외교현안 발생시 분석 등으로 나누어 제시할 수 있다.

자동화된 분석 결과를 최고 정책결정자가 직접 본다는 것이 아니라, 실무진 차원에서 일단이상에서 제시한 과정의 빅데이터 분석을 수행하고, 그 중에서도 의미 있는 것을 일정한 단위로 실·국장 수준에서 보고받는 방법을 채택하는 것이 유용할 것이다. 다시 말해 최고 정책결정자가 분석의 원자료 전체를 일일이 살펴볼 수 없는 상황에서 컴퓨터가 패턴화한 것을 실무진들이 보고 난 후 의미 있다고 여겨 정리해 놓은 것을, 일종의 보관창고에 넣어 놓고, 필요에 따라서 꺼내서 재조합하여 의사결정에 활용하는 작업흐름(work flow) 체계를 마련하는 것이 필요하다. 현재 빅데이터의 도입 추세로 미루어보건대, 빅데이터를 기반으로 하는 의사결정 도구들의 도입은 길지 않은 기간 내에 상당히 보편화될 것으로 보이며 특정 부분에서는 가속화될 것으로 예견되고 있다.

IV. 맺음말

향후 빅데이터를 활용한 외교안보 정책을 추진하기 위해서는 현재 진행하고 있는 분석 프레임워크 개발 연구를 바탕으로 하여, 향후 시뮬레이션 작업과 시뮬레이션을 통한 모듈 개발 연구를 수행할 필요가 있다. 이를 바탕으로 빅데이터 분석 시스템의 본격적인 가동을 가능케 할 수 있을 것으로 기대한다. 이를 위한 소요 예산의 확보 및 외교안보 관련 부처의 조직정비가 필요함은 물론이다. 온라인 전담조직을 신설하되, 애초 모니터링 목적이 아닌 인텔리전스 목적으로 설립할 필요가 있다. 첨예한 외교 현안에 대해 질문을 던지고, 데이터 분석을 통해 답을 제시하는 조직을 마련할 필요가 있다. 이를 바탕으로 외교안보 정책 담당 부처내와 학계·싱크탱크 및 업계의 협업체계를 구축하여야 할 것이다. 이를 바탕으로 시범분석과 DB구축을 수행하고 더 나아가 제도화된 사업으로 운영할 필요가 있다. 외교안보 빅데이터 분석의 전담조직이 구축되고, 학계·산업계와 협업을 구축해서, 한 개의 사안에 대해서 시범적으로 분석을 해보고 DB를 구축한 다음, 계속해서 분석사례를 축적해 나가면서 가이드 라인을 찾을 수 있을 것이다.

특히 외교안보 담당 정부부처, 외교안보 연구기관, 빅데이터 분석업체의 효과적인 협업체계 구축이 중요하다. 협업 시스템을 구축하는 것이 중요하다는 것은 막연하게 협업하는 것이 아니라 구체적으로 제도화된 시스템을 구축해야 한다는 의미이다. 다시 말해, 빅데이터를 활용한 외교안보정책의 수행을 위해서 필요한 공공 연구기관과 민간 연구소 및 싱크탱크의 협업체계 구축 필요성, 그리고 더 나아가 이를 실제 외교안보 정책에 활용하는 정책 지식 협업체계 구축에 대한 검토도 필요하다. 빅데이터 분석과 외교안보 정책 관련 작업의



학제적, 학술-실무 연계적 성격 등을 고려할 때 기초연구 → 구조화 → 협업 → 구체적 정책 타깃팅 등에 이르는 전 과정에 대한 디자인 설계의 마인드가 필요하다. 더 나아가 빅데이터 분석기술과 DB구축을 밑받침으로 하여, 정부 부처 간 및 업체와 싱크탱크 간 협업의 공동플랫폼을 구축할 필요가 있다. 외교안보 관련 정부부처와 외교안보 연구기관은 업체를 통해서 빅데이터 분석업체와의 적절한 분업을 통해서 분석을 수행하며, 이러한 과정에서 일종의 ‘감독’의 역할을 수행할 수 있을 것이다.

〈참고문헌〉

- 강선주. 2014. “빅데이터 구축 현황과 외교안보적 활용 방향.” 『주요국제문제분석』 No.2014-10. 국립외교원 외교안보연구소.
- 김도훈. 2017. “빅데이터를 활용한 정책공공외교: 주요 어젠다.” 서울대학교 국제문제연구소 세미나 발표자료. 2017.4.14
- 김상배. 2004. “지식/네트워크의 국가전략: 외교분야를 중심으로.” 『국가전략』 10(1), pp.167-194.
- 김상배. 2015. “빅데이터의 국가전략: 21세기 신흥권력 경쟁의 개념적 성찰.” 『국가전략』 21(3), pp.5-35.
- 매일경제 기획팀 · 서울대빅데이터센터. 2014. 『빅데이터 세상』 매일경제신문사.
- 박종희. 2015. “21세기 외교에서 빅데이터 활용 가능성에 대한 탐구.” 한국국제정치학회 하계학술대회 발표문.
- 배영자. 2015. “빅데이터와 외교: 미국과 한국 사례 연구.” 한국국제정치학회 하계학술대회 발표문.
- 송영조. 2015. “데이터 기반의 미래전략기구와 정책적 함의: 싱가포르 RAHS사례 중심으로.” 빅데이터와 중장기 외교전략 세미나 발표문. 서울대학교. 4월 16일.
- 쉴버거, 빅토르 마이어 · 케네스 쿠키어. 2013. 『빅데이터가 만드는 세상』 21세기북스.
- 이상원. 2017. “빅데이터를 활용한 정책공공외교.” 서울대학교 국제문제연구소 세미나 발표자료. 2017.4. 21
- 이승주. 2015. “빅데이터와 외교정책: 영국과 싱가포르의 사례를 중심으로.” 한국국제정치학회 하계학술대회 발표문.
- 이영주. 2017. “빅데이터를 활용한 미래예측 및 공공정책 지원 사례” 서울대학교 국제문제연구소 세미나 발표자료. 2017.5.11
- 정용찬. 2013. 『빅데이터』 커뮤니케이션북스.
- 조현석 편. 2013. 『빅데이터와 위험정보사회』 커뮤니케이션북스.
- 채승병. 2015. “빅데이터 비즈니스의 동향과 쟁점.” 빅데이터와 중장기 외교전략 세미나 발표문. 서울대학교. 4월 9일.
- Esposti, Sara Degli. 2014. “When Big Data Meets Dataveillance: the Hidden



Side of Analytics.” *Surveillance & Society*, 12(2), pp.209–225.

Hansen, Hans Krause, and Tony Porter. 2015. “What do Big Data do in Transnational Governance?” Paper Presented at the International Studies Association Meetings, New Orleans, February 21, 2015.

Roderick, Leanne. 2014. “Discipline and Power in the Digital Age: The Case of the US Consumer Data Broker Industry.” *Critical Sociology*. 40(5), pp.729–746.

Skelly, James and Christian Eichenmüller. 2015. “Surveillance, ‘Big Data,’ and the Future of Global Politics.” Paper Presented at the International Studies Association Meetings, New Orleans, February 21, 2015.

Chandler, David. 2015. “A World without Causation: Big Data and the Coming of Age of Posthumanism.” *Millennium*, 43(3), pp.833–851.

Zwitter, Andrej. 2015. “Big Data and International Relations.” *Ethics & International Affairs*, 29(4), pp.377–389.