

소셜 빅데이터를 활용한 미래신호 예측: 보건복지 주요 정책과 이슈를 중심으로



Using Social Big Data Predictive
Future Signal: With Special Reference to the
Major Policy Issues of Health and Welfare

송태민 | 한국보건사회연구원 선임연구위원

보건복지 정책과 이슈의 미래신호 분석에서 미래세대육성과 개인정보는 강신호이면서 높은 증가율을 보이고 있어, 미래세대육성 정책에 포함된 아동학대 문제의 해결과 개인정보보호와 관련된 제도 개선에 대한 논의가 지속적으로 이루어져야 할 것이다. {건강보험, 중증질환}이 동시에 언급된 문서에 대한 찬성 비율이 높은 것으로 나타났는데 건강보험 혜택 확대로 4대중증질환의 보장성이 강화되어 국민의 의료비 부담을 줄인 것이 정부의 정책에 대한 좋은 평가로 나타났다. 소셜 빅데이터를 활용함으로써 정부의 정책과 새로운 기술에 대한 미래신호를 사전에 예측하여 대비할 수 있으며, 사회적 위기 상황에 대한 모니터링과 예측으로 위험에 대한 사전 대응 체계를 구축할 수 있다. 빅데이터를 분석하여 인과성과 연관성을 발견하고 미래를 예측하기 위해서는 데이터사이언티스트 양성을 위한 정부 차원의 노력이 필요하다.

1. 들어가며¹⁾

미래변화의 트렌드를 파악하고 미래의 핵심 기술을 선별하기 위하여 주요 선진국들은 주기

적으로 국가의 미래 트렌드를 분석해 결과를 발표하고 있다.²⁾ 그동안 미래 트렌드를 예측하기 위한 다양한 연구가 시도되어 왔으나 대부분 전문가의 지식과 의견에 따라 미래를 전망하는 방

1) 이 논문은 미래창조과학부·정보통신기술진흥센터의 연구·개발(R&D) 프로그램[R7117-16-0219, 점진적 기계학습 기반 자가진화(Self-Evolving) 에이전트 시뮬레이션을 이용한 사회변화 예측분석 기술 개발]의 지원을 받아 작성되었음.

2) 정근하(2010). 텍스트마이닝과 네트워크 분석을 활용한 미래예측 방법 연구. 한국과학기술기획평가원. p.6.

법을 사용하여 왔다. 최근 소셜네트워크서비스(SNS)를 비롯한 온라인 채널에서 생산되는 텍스트 형태의 비정형 데이터가 실제 경제 및 사회에 미치는 영향력이 매우 높아짐에 따라 소셜 빅데이터를 활용한 미래예측 연구가 진행되고 있으나 수집 기술과 분석 기술의 어려움으로 활발히 확산되지 못하고 있는 실정이다.

현 정부는 국민이 행복한 사회를 이루기 위한 사회보장 정책의 방향으로 '생애주기별 맞춤형 복지'를 제시하고, 이를 실현하기 위한 맞춤형 복지정책을 도입 및 확대하고 있다. 2016년도에는 국민이 체감하는 맞춤형 복지 확산을 목표로 맞춤형 기초생활보호제도 정착, 4대중증질환 등의료보장 지속, 맞춤형 보육 개편, 기초연금 및 장기요양 지원 확대, 복지 안내 강화, 취약계층 필수 서비스 지속, 국민 체감형 원격의료 확산, 신약 개발 및 제약산업 육성 등을 중점적으로 추진해 왔다. 정부가 보건복지 정책을 성공적으로 추진하여 예상하는 성과를 얻기 위해서는 다양한 보건복지 욕구와 이해집단과의 갈등을 최소화하기 위한 정책 동향 및 수요를 예측하여 적시에 대응할 수 있는 체계를 구축해야 한다.

본 연구는 우리나라에서 수집 가능한 모든 온라인 채널에서 언급된 보건복지 관련 문서를 수집하여 주제 분석과 감성 분석을 통하여 보건복

지 주요 키워드를 분류하고 보건복지와 관련하여 나타나는 주요 정책과 이슈에 대한 미래신호를 탐지하여 예측 모형을 제시하고자 한다.

2. 소셜 빅데이터 기반 미래예측 방법론

오늘날 미래의 환경 변화를 감지하기 위한 다양한 연구가 시도되고 있으며, 여러 연구 중에서 가장 많이 주목받고 있는 것은 미래의 변화를 예감할 수 있는 약신호(weak signal)를 탐지하는 것이다.³⁾ 약신호는 '미래에 가능한 변화의 징후'⁴⁾로 이는 시간이 흐르면서 강신호(strong signal)로, 강신호는 다시 트렌드(trend)나 메가트렌드(mega trend)로 발전할 수 있다. Hiltunen⁵⁾은 약신호에 대해 미래신호(future signal)라는 개념을 이용하여 미래신호를 신호(signal), 이슈(issue), 이해(interpretation)와 같이 3차원의 공간으로 설명하였다. Yoon은 웹 뉴스의 문서를 수집하여 텍스트마이닝 분석을 통해 생성된 단어빈도(Term Frequency, TF)와 문서빈도(Document Frequency, DF)를 Hiltunen의 신호, 이슈와 각각 연계하였다. Yoon은 단어빈도, 문서빈도, 발생빈도 증가율을 이용하여 KEM(Keyword Emergence Map)

3) Yoon, J., Detecting weak signals for long-term business opportunities using text mining of Web news, Journal Expert Systems with Applications 39(16), 2012, pp.12543-12550.; 박찬국&김현제(2015). 사물인터넷을 통한 에너지 신산업 발전 방향 연구-텍스트마이닝을 이용한 미래 신호 탐색. 에너지경제연구원.

4) Ansoff, H.I., Managing strategic surprise by response to weak signals, Californian Management Review 18(2), 1975, pp.21-33.

5) Hiltunen, E., The future sign and its three dimensions, Futures 40, 2008, pp.247-260.

과 KIM(Keyword Issue Map)의 키워드 포트폴리오를 작성하고 이를 이용하여 약신호를 선별하였다. KEM은 가시성을 보여 주는 것으로 DoV(Degree of Visibility)를 산출하고, KIM은 확산 정도를 보여 주는 것으로 DoD(Degree of Diffusion)를 산출할 수 있다.

$$DoV_{ij} = \left(\frac{TF_{ij}}{NN_j} \right) \times \{ 1 - tw \times (n - j) \}$$

$$DoD_{ij} = \left(\frac{DF_{ij}}{NN_j} \right) \times \{ 1 - tw \times (n - j) \}$$

여기서 NN 은 전체 문서 수를 의미하고, TF 는 단어빈도, DF 는 문서빈도, tw 는 시간가중치(본 연구에서 시간가중치는 0.05를 적용), n 은 전체 시간 구간, j 는 시점을 의미한다.

3. 소셜 빅데이터 기반 미래신호 탐색

가. 연구 대상

본 연구는 국내의 온라인 뉴스 사이트, 블로그, 카페, 소셜네트워크서비스, 게시판 등 인터넷을 통해 수집된 소셜 빅데이터를 대상으로 하였다. 본 분석에서는 온라인 뉴스사이트 149개, 블로그(네이버, 네이트, 다음, 티스토리) 4개, 카페

(네이버, 다음) 2개, SNS(트위터) 1개, 게시판(네이버 지식인, 네이트 지식, 네이트 톨, 네이트 판 등) 15개 등 총 171개의 온라인 채널을 통해 수집 가능한 텍스트 기반의 웹문서(버즈)를 소셜 빅데이터로 정의하였다. 보건복지 관련 토픽의 수집은 2016. 1. 1.~2016. 3. 31. 해당 채널에서 요일별, 주말, 휴일을 고려하지 않고 매 시간 단위로 수집하였으며, 수집된 총 201,849건(1월: 87,567건, 2월: 65,278건, 3월: 49,004건)의 텍스트(Text) 문서를 본 연구의 분석에 포함시켰다. 본 연구를 위한 소셜 빅데이터의 수집⁶⁾에는 크롤러(Crawler)를 사용하였고, 토픽의 분류는 주제분석기법을 사용하였다. 보건복지 토픽은 모든 관련 문서를 수집하기 위해 ‘보건’, ‘복지’, 그리고 ‘보건복지’를 사용하였다.

나. 단어 및 문서 빈도 분석⁷⁾

온라인 채널에서 수집된 텍스트 형태의 문서를 분석하려면 먼저 텍스트마이닝을 통하여 단어빈도와 문서빈도를 산출⁸⁾해야 한다. 그리고 중요한 정보의 추출을 위해서 TF-IDF(Term Frequency - Inverse Document Frequency) 방법을 사용하고 있다.

Spärck⁹⁾는 희귀한 단어일수록 더 높은 가중치를 부여하기 위해서 역문서빈도(Inverse

6) 본 연구를 위한 소셜 빅데이터의 수집 및 토픽 분류는 ‘(주)SK텔레콤 스마트인사이트’에서 수행함.

7) 본 단어빈도 및 문서빈도의 분석 절차는 박찬국&김현제(2015)의 분석 방법을 참고한 것임을 밝힘.

8) 단어빈도(Term Frequency, TF)는 각 문서에서 단어별 출현 빈도를 산출한 후, 문서별 출현 빈도를 합산하여 산출할 수 있다. 문서빈도(Document Frequency, DF)는 특정 단어가 들어가는 문서의 수를 나타낸다.

9) Spärck Jones, K. (1972). “A Statistical Interpretation of Term Specificity and Its Application in Retrieval”. Journal of Documentation 28: 11-21.

Document Frequency, $IDF_j = \log_{10}(\frac{N}{DF_j})$)를 제안하였다. 단어빈도 분석에서 희귀한 단어일수록 더 높은 가중치를 부여할 필요가 있다면 단어빈도와 역문서빈도를 결합하여 'TF-IDF = $TF_i \times IDF_j$ '를 산출하여 가중치(단어의 중요도 지수)를 적용한다.

상기 분석 방법론에 따라 단어빈도, 문서빈도, 단어의 중요도 지수를 고려한 문서빈도의 분석을 통하여 보건복지 관련 정책과 주요 이슈에 대한 인식 변화를 살펴보았다(표 1). 단어빈도에서는 일자리, 증세, 세금, 복지급여, 결혼출산, 건강증진, 치료 등의 순위로 나타나고 있는데 정책은 일자리, 복지급여, 결혼출산이 우선이고 주요 이슈는 증세, 세금, 치료가 우선임을 보여 준

다. 문서빈도는 단어빈도와 비슷한 추이를 나타내고 있으나 결혼출산이 단어빈도에서는 5위인 반면 문서빈도에서는 7위로 나타났다. 키워드의 중요성을 나타내는 단어빈도에서는 결혼출산이 중요하나 주제의 확산을 나타내는 문서빈도에서는 순위가 다소 떨어져 결혼출산 정책의 확산을 위한 노력이 필요할 것으로 보인다. 중요도 지수를 고려한 단어빈도로 파악한 정책은 일자리, 복지급여, 결혼출산이 우선이고 주요 이슈는 증세, 세금순으로 나타났다. 키워드의 월별 순위 변화를 살펴보면 2016년 2월까지 증세, 일자리, 세금, 복지급여, 치료가 중요한 키워드로 나타나다가 3월에는 건강증진이 강조되기 시작하여 건강에 대한 관심이 확산되고 있는 것으로 나타났다.

표 1. 온라인 채널의 보건복지 정책, 이슈의 키워드 분석

순위	TF		DF		TF-IDF	
	키워드	빈도	키워드	빈도	키워드	빈도
1	일자리	8,212	증세	8,059	일자리	6,328
2	증세	8,059	일자리	7,459	증세	5,940
3	세금	5,339	세금	5,339	복지급여	4,955
4	복지급여	4,520	복지급여	3,524	세금	4,890
5	결혼출산	3,419	건강증진	3,352	결혼출산	4,267
6	건강증진	3,352	치료	2,637	건강증진	3,748
7	치료	2,938	결혼출산	2,485	치료	3,591
8	건강보험	1,307	무상정책	1,116	건강보험	2,114
9	무상정책	1,156	건강보험	1,062	무상정책	1,845
10	기초연금	922	기초연금	922	기초연금	1,548
11	개인정보	872	개인정보	872	개인정보	1,485
12	미래세대육성	817	미래세대육성	817	미래세대육성	1,414
13	의료비	754	의료비	754	의료비	1,332

(표 계속)

순위	TF		DF		TF-IDF	
	키워드	빈도	키워드	빈도	키워드	빈도
14	국민연금	738	국민연금	738	국민연금	1,310
15	의료민영화	686	의료민영화	662	보건산업	1,263
16	보건산업	672	자살	634	의료민영화	1,250
17	자살	634	보건산업	581	자살	1,167
18	부동산	483	부동산	483	부동산	946
19	등록금	471	등록금	459	보육	939
20	보육	460	보육	400	등록금	933
21	담배	383	담배	383	담배	789
22	가족친화	348	양극화	306	가족친화	761
23	양극화	306	가족친화	287	양극화	660
24	원격의료	237	중증질환	223	원격의료	546
25	중증질환	233	원격의료	219	중증질환	535
26	환자안전	229	환자안전	212	환자안전	531
합계		47,547	합계	43,985	합계	55,087

4. 보건복지 관련 키워드의 미래신호 탐색¹⁰⁾

미래신호 탐지 방법론에 따라 분석한 결과는 <표 2, 표 3>과 같다. 보건복지 관련(정책, 이슈) 키워드에 대한 DoV 증가율과 평균단어빈도를 산출한 결과 일자리와 복지급여는 높은 빈도를 보이고 있으나 DoV 증가율은 중앙값보다 낮게 나타나 시간이 갈수록 신호가 약해지는 것으로 나타났다. 결혼출산, 건강증진 정책의 평균단어빈도는 높게 나타났으며, DoV 증가율은 중앙값보다 높게 나타나 시간이 갈수록 빠르게 신호가 강해지는 것으로 나타났다.

미래신호 탐색을 위해 DoV의 평균단어빈도와

DoD의 평균문서빈도를 X축으로 설정하고 DoV와 DoD의 평균증가율을 Y축으로 설정한 후 각 값의 중앙값을 사분면으로 나누면 2사분면에 해당하는 영역의 키워드는 약신호가 되고 1사분면에 해당하는 키워드는 강신호가 된다. 빈도 수 측면에서 볼 때 상위 10위에 DoV는 일자리, 증세, 세금, 복지급여, 결혼출산, 건강증진, 치료, 건강보험, 무상정책, 기초연금의 순으로 포함되었고 DoD에는 증세, 일자리, 세금, 복지급여, 건강증진, 치료, 결혼출산, 무상정책, 건강보험, 기초연금의 순으로 포함되었다. DoV의 평균증가율의 중앙값(0.23)보다 높은 증가율을 보이는 키워드는 결혼출산, 건강증진, 건강보험으로 나타났으

10) 본 미래신호 탐색 절차는 박찬국&김현제(2015)의 분석 방법을 참고한 것임을 밝힘.

며 DoD의 평균증가율의 중앙값(0.23)보다 높은 증가율을 보이는 키워드는 건강증진, 건강보험으로 나타났다. 특히 결혼출산의 DoV증가율은

중앙값보다 높은 반면 DoD의 증가율은 중앙값보다 낮게 나타나 결혼출산 정책의 확산을 위한 방안이 필요할 것으로 보인다.

표 2. 보건복지 정책, 이슈의 DoV 평균증가율과 평균단어빈도

키워드	DoV			평균증가율	평균단어빈도
	1월	2월	3월		
일자리	0.142	0.168	0.186	0.147	2737
증세	0.158	0.237	0.063	-0.116	2686
세금	0.136	0.098	0.077	-0.246	1780
복지급여	0.11	0.067	0.094	0.008	1507
결혼출산	0.064	0.049	0.097	0.369	1140
건강증진	0.062	0.051	0.094	0.339	1117
치료	0.045	0.07	0.061	0.208	979
건강보험	0.026	0.02	0.034	0.245	436
무상정책	0.024	0.021	0.024	0.014	385
기초연금	0.013	0.02	0.023	0.33	307
개인정보	0.005	0.04	0.004	3.394	291
미래세대육성	0.01	0.005	0.04	3.498	272
의료비	0.015	0.01	0.022	0.415	251
국민연금	0.007	0.014	0.026	0.886	246
의료민영화	0.003	0.004	0.04	5.242	229
보건산업	0.011	0.013	0.017	0.216	224
자살	0.006	0.005	0.03	2.167	211
부동산	0.009	0.012	0.007	-0.034	161
등록금	0.011	0.007	0.011	0.1	157
보육	0.008	0.012	0.006	0.015	153
담배	0.005	0.005	0.014	0.827	128
가족친화	0.004	0.004	0.014	1.275	116
양극화	0.01	0.003	0.004	-0.226	102
원격의료	0.005	0.005	0.004	-0.103	79
중증질환	0.007	0.004	0.003	-0.307	78
환자안전	0.002	0.008	0.003	1.062	76
중앙값				.23	249

표 3. 보건복지 정책, 이슈의 DoD 평균증가율과 평균문서빈도

키워드	DoD			평균증가율	평균문서빈도
	1월	2월	3월		
증세	0.172	0.251	0.07	-0.131	2686
일자리	0.138	0.163	0.186	0.159	2486
세금	0.148	0.104	0.085	-0.238	1780
복지급여	0.092	0.055	0.082	0.047	1175
건강증진	0.067	0.054	0.104	0.369	1117
치료	0.043	0.068	0.06	0.221	879
결혼출산	0.054	0.042	0.069	0.224	828
무상정책	0.026	0.021	0.026	0.015	372
건강보험	0.022	0.017	0.032	0.309	354
기초연금	0.014	0.021	0.026	0.337	307
개인정보	0.005	0.042	0.005	3.278	291
미래세대육성	0.011	0.005	0.044	3.683	272
의료비	0.016	0.011	0.024	0.453	251
국민연금	0.008	0.015	0.028	0.902	246
의료민영화	0.003	0.003	0.044	6.347	221
자살	0.007	0.006	0.033	2.279	211
보건산업	0.01	0.013	0.015	0.229	194
부동산	0.01	0.013	0.008	-0.04	161
등록금	0.011	0.007	0.012	0.12	153
보육	0.008	0.012	0.005	0.015	133
담배	0.006	0.005	0.015	0.875	128
양극화	0.011	0.003	0.004	-0.203	102
가족친화	0.004	0.004	0.012	1.047	96
중증질환	0.007	0.003	0.003	-0.265	74
원격의료	0.005	0.005	0.004	-0.119	73
환자안전	0.002	0.008	0.003	1.27	71
중앙값				.23	249

[그림 1], [그림 2]와 같이 보건복지 관련 주요 키워드인 결혼출산은 KEM에서 강신호로 나타난 반면 KIM에서는 강하진 않지만 잘 알려진 신호로 나타났다. KEM과 KIM에 공통적으로 나타나는 강신호(1사분면)에는 미래세대육성, 개인정보,

의료비, 기초연금, 건강보험, 치료, 건강증진이 포함되었고 약신호(2사분면)에는 의료민영화, 자살, 환자안전, 가족친화, 담배, 보건산업, 국민연금이 포함된 것으로 나타났다. KEM과 KIM의 4사분면에 공통적으로 나타난 강하진 않지만 잘 알려진

그림 1. 보건복지 정책, 이슈의 KEM(Keyword Emergence Map)

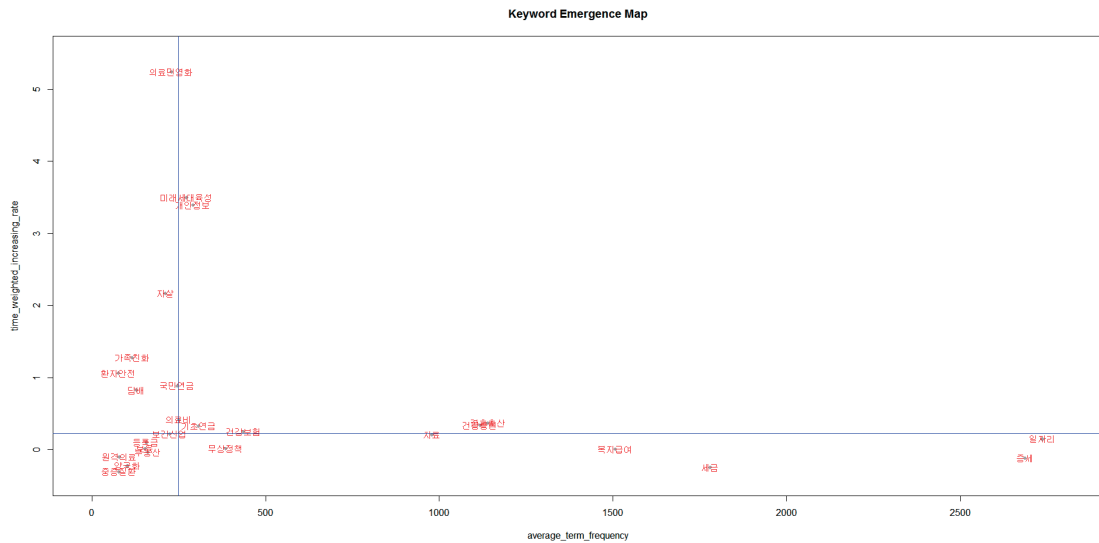
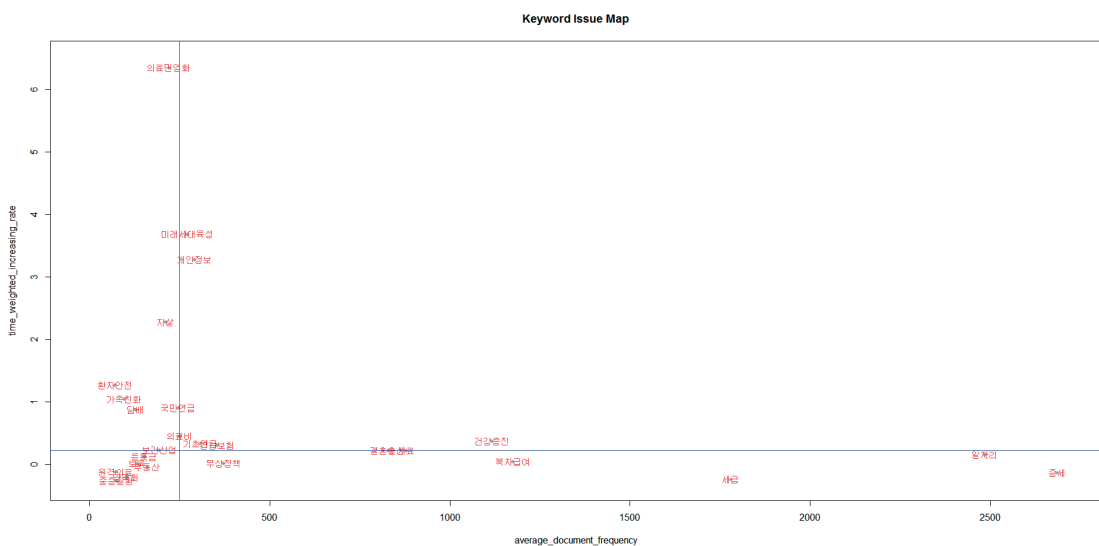


그림 2. 보건복지 정책, 이슈의 KIM(Keyword Issue Map)



신호는 무상정책, 복지급여, 세금, 일자리, 증세로 나타났으며 KEM과 KIM의 3사분면에 공통적으로 나타난 잠재신호는 등록금, 보육, 부동산, 원격의료, 양극화, 중증질환으로 나타났다(표 4).

표 4. 보건복지 정책, 이슈의 미래신호

구분	잠재신호 (Latent signal)	약신호 (Weak Signal)	강신호 (Strong signal)	강하진 않지만 잘 알려진 신호 (Not strong but well known signal)
KEM	등록금, 보육, 부동산, 원격의료, 양극화, 중증질환	의료민영화, 자살, 가족친화, 환자안전, 담배, 보건산업, 국민연금	미래세대육성, 개인정보, 의료비, 기초연금, 건강보험, 결혼출산, 치료, 건강증진,	무상정책, 복지급여, 세금, 일자리, 증세
KIM	등록금, 보육, 부동산, 원격의료, 양극화, 중증질환	의료민영화, 자살, 환자안전, 가족친화, 담배, 보건산업, 국민연금	미래세대육성, 개인정보, 의료비, 기초연금, 건강보험, 치료, 건강증진	무상정책, 결혼출산, 복지급여, 세금, 일자리, 증세
주요 신호	등록금, 보육, 부동산, 원격의료, 양극화, 중증질환	의료민영화, 자살, 환자안전, 가족친화, 담배, 보건산업, 국민연금	미래세대육성, 개인정보, 의료비, 기초연금, 건강보험, 치료, 건강증진	무상정책, 복지급여, 세금, 일자리, 증세

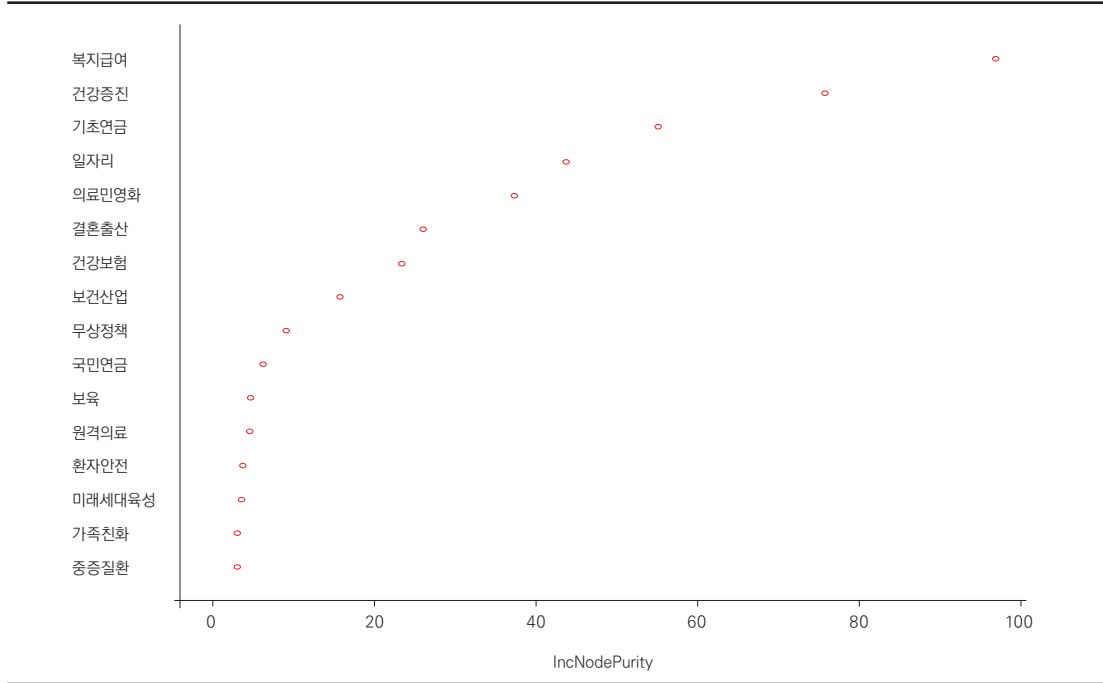
5. 보건복지 정책 미래신호 예측

가. 랜덤포레스트 분석을 통한 주요 보건복지 정책 예측

본 연구의 랜덤포레스트(Random Forest) 분석을 활용하여 보건복지 수요(태도)에 영향을 주

는 주요 정책을 살펴보면 [그림 3]과 같다. 보건복지 수요(찬성, 반대)에 가장 큰 영향을 미치는 (연관성이 높은) 정책은 ‘복지급여’ 정책으로 나타났다으며 뒤를 이어 건강증진, 기초연금, 일자리, 의료민영화, 결혼출산, 건강보험, 보건산업 정책 등의 순으로 나타났다.

그림 3. 랜덤포레스트 모델의 보건복지 주요 정책의 중요도



랜덤포레스트의 중요도로 나타난 정책들이 보건복지 수요에 미치는 영향을 로지스틱 회귀분석으로 살펴본 결과 기초노령연금, 의료민영화는 반대의 확률이 높으며 그 외 국민연금(p<.1),

보육, 출산양육, 가족친화, 건강보험, 원격의료 정책 등은 찬성의 확률이 높은 것으로 나타났다 (표 5).

표 5. 보건복지 수요에 영향을 주는 주요 정책 요인

정책	찬성 ¹⁾			
	b [†]	S.E. [‡]	OR [§]	P
국민연금	.254	.153	1.289	.098
기초연금	-1.483	.100	.227	.000
보육	1.440	.318	4.221	.000
결혼출산	.629	.070	1.875	.000
가족친화	1.119	.300	3.061	.000

(표 계속)

정책	찬성 ¹⁾			
	b [†]	S.E. [‡]	OR [§]	P
미래세대육성	.796	.212	2.216	.000
무상정책	.405	.120	1.499	.001
의료민영화	-1.634	.121	.195	.000
건강보험	1.180	.131	3.256	.000
원격의료	1.133	.316	3.106	.000
중증질환	1.037	.342	2.821	.002
환자안전	1.303	.351	3.681	.000
보건산업	1.625	.232	5.079	.000
복지급여	1.225	.068	3.405	.000
건강증진	1.041	.065	2.831	.000
일자리	.501	.045	1.650	.000

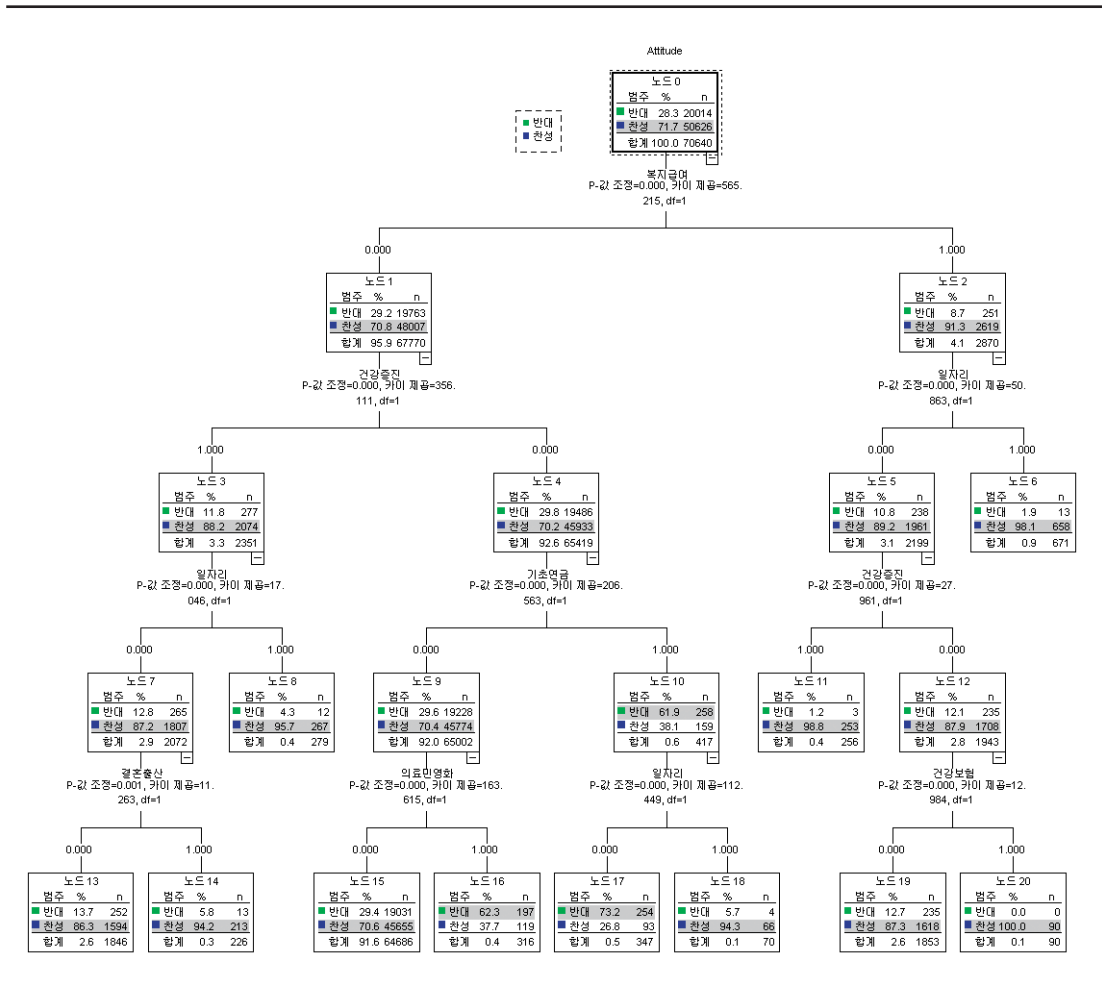
주: 1) 기준범주는 반대, [†] Standardized coefficients, [‡] Standard error, [§] odds ratio

나. 의사결정나무 분석을 통한 보건복지 정책 수요 예측

보건복지 정책 수요 예측 모형에 대한 의사결정나무는 [그림 4]와 같다. 나무 구조의 최상위에 있는 뿌리나무는 예측변수(독립변수)가 투입되지 않은 종속변수의 빈도를 나타낸다. 뿌리 마디의 보건복지에 대한 감정의 비율을 보면 보건복지에 대한 찬성은 71.7%, 반대는 28.3%로 나

타났다. 뿌리 마디 하단의 가장 상위에 위치하는 정책이 종속변수에 대한 영향력이 가장 높은(관련성이 깊은) 정책이므로 ‘복지급여’ 정책이 가장 영향력이 큰 것으로 나타났다. 즉, 온라인 문서에 ‘복지급여’ 정책이 있는 경우 찬성은 이전의 71.7%에서 91.3%로 증가하였다. ‘복지급여’ 정책과 ‘일자리’ 정책이 있는 경우 찬성은 이전의 91.3%에서 98.1%로 증가하였다.

그림 4. 보건복지 정책 수요 예측의 의사결정나무 모형



다. 연관분석을 통한 보건복지 정책 수요 예측

소셜 빅데이터 분석에서 연관분석은 하나의 온라인 문서에 포함된 둘 이상의 단어들에 대한 상호 관련성을 발견하는 것이다. 본 연구에서는 <표 6>과 같이 하나의 문서에 나타난 정책 요인의 수요에 대한 연관규칙을 분석하였다. {의

료민영화} ⇒ {반대} 두 변인의 연관성은 지지도 0.003, 신뢰도는 0.579, 향상도는 2.044로 나타났다. 이는 온라인 문서에서 ‘의료민영화’ 정책이 언급되면 반대의 확률이 57.9%이며 ‘의료민영화’ 정책이 언급되지 않은 문서보다 반대의 확률이 약 2.04배 높아지는 것을 나타낸다. {건강보

힘, 중증질환} ⇒ {찬성}으로 세 변인의 향상도는 정책이 언급되지 않은 문서보다 찬성할 확률이 1.40으로 온라인 문서에서 건강보험과 중증질환 1.40배 높은 것으로 나타났다.

표 6. 보건복지 주요 정책의 연관규칙

규칙	지지도	신뢰도	향상도
{의료민영화} ⇒ {반대}	0.002845413	0.57925072	2.0444824
{기초연금} ⇒ {반대}	0.003680634	0.44444444	1.5686797
{건강보험, 중증질환} ⇒ {찬성}	0.001670442	1.00000000	1.3953305
{출산양육, 건강증진} ⇒ {찬성}	0.001528879	1.00000000	1.3953305
{출산양육, 복지급여} ⇒ {찬성}	0.002180068	1.00000000	1.3953305
{보건산업, 복지급여} ⇒ {찬성}	0.001005096	1.00000000	1.3953305
{국민연금, 복지급여} ⇒ {찬성}	0.002109287	1.00000000	1.3953305
{건강보험, 복지급여} ⇒ {찬성}	0.002831257	1.00000000	1.3953305
{출산양육, 복지급여, 건강증진} ⇒ {찬성}	0.001090034	1.00000000	1.3953305
{출산양육, 건강증진, 일자리} ⇒ {찬성}	0.001005096	1.00000000	1.3953305
{출산양육, 복지급여, 일자리} ⇒ {찬성}	0.001373160	1.00000000	1.3953305
{국민연금, 건강보험, 복지급여} ⇒ {찬성}	0.001061721	1.00000000	1.3953305
{국민연금, 복지급여, 일자리} ⇒ {찬성}	0.001160815	1.00000000	1.3953305
{건강보험, 복지급여, 건강증진} ⇒ {찬성}	0.001061721	1.00000000	1.3953305
{건강보험, 복지급여, 일자리} ⇒ {찬성}	0.001443941	1.00000000	1.3953305
{중증질환, 복지급여} ⇒ {찬성}	0.001613817	0.99130435	1.3831972
{기초연금, 복지급여, 일자리} ⇒ {찬성}	0.001359003	0.98969072	1.3809456
{복지급여, 건강증진, 일자리} ⇒ {찬성}	0.002548131	0.98901099	1.3799972
{복지급여, 건강증진} ⇒ {찬성}	0.006129672	0.98858447	1.3794020
{국민연금, 건강보험, 일자리} ⇒ {찬성}	0.001160815	0.98795181	1.3785193

6. 나가며

보건복지 정책과 이슈의 미래신호를 예측하기 위해서는 연도별 온라인 문서에 대한 시계열 정보가 필요한데 본 연구는 2016. 1.~2016. 3. 3개월간의 정보를 수집하여 미래신호 예측 방법을 우선적으로 적용해 보았다. 본 연구의 정보 분석 기간의 한계에도 불구하고 보건복지 정책과 이슈에 대한 미래신호 예측 결과를 살펴보면 다음과 같다. 첫째, 본 연구의 보건복지 정책과 이슈의 미래신호 분석에서 미래세대육성, 기초연금, 건강보험, 건강증진, 개인정보, 의료비, 치료가 강신호로 분류되어 미래세대육성 및 건강증진과 관련된 의료비와 치료 등이 강조되고 있는 것으로 나타났다. 특히, 미래세대육성과 개인정보는 강신호이면서 높은 증가율을 보이고 있어 미래세대육성 정책에 포함된 아동학대 문제의 해결과 개인정보보호와 관련된 제도 개선에 대한 논의가 지속적으로 이루어져야 할 것이다. 둘째, 의료민영화, 자살, 환자안전, 가족친화, 담배, 보건산업, 국민연금은 약신호로 분류되었다. 특히 약신호인 의료민영화와 자살은 높은 증가율을 보이고 있는데 이들 키워드는 시간이 지나면 강신호로 발전할 수 있기 때문에 대응책이 마련되어야 할 것으로 본다. 보건복지 정책의 미래신호 예측에서 중요한 정책이면서 찬성하는 정책은 복지급여, 건강증진, 일자리, 결혼출산, 건강보험, 보건

산업 등의 순으로 나타났다. 특히 복지급여와 일자리가 동시에 언급된 문서에 대한 찬성이 매우 높은 것으로 나타나 능동적 복지체계 구축을 통한 일자리 창출이 필요할 것으로 본다. 이는 정책의 연관분석 결과와 같이 기초연금 정책만 언급된 문서는 반대의 확률이 높은 것으로 나타났으나 기초연금, 복지급여, 일자리 정책이 동시에 언급된 문서는 찬성의 확률이 높은 것으로 나타났다. 이는 노인의 능동적 자활과 근로를 통한 복지체계 구축에 대한 국민의 요구가 높은 것으로 보인다. 건강보험, 중증질환이 동시에 언급된 문서는 찬성의 비율이 높은 것으로 나타나 건강보험 혜택 확대로 4대중증질환의 보장성이 강화되어 국민의 의료비 부담을 줄임으로써 정부의 정책이 좋은 평가를 받은 것으로 보인다.

소셜 빅데이터는 다양한 분야에 활용할 수 있다. 첫째, 본 연구의 적용과 같이 정부의 정책과 새로운 기술에 대한 미래신호를 사전에 예측하여 대비할 수 있다. 둘째, 조사를 통한 기존의 정보 수집 체계의 한계를 보완할 수 있는 새로운 자료 수집 방법으로 활용할 수 있다. 셋째, 사회적 위기 상황에 대한 모니터링과 예측으로 위험에 대한 사전 대응체계를 구축할 수 있다.

끝으로 빅데이터를 분석하여 인과성과 연관성을 발견하여 미래를 예측하기 위해서는 데이터 사이언티스트 양성을 위한 정부 차원의 노력이 필요하다. ■