

사행산업 관련 뉴스의 빅데이터 분석을 통한 정책 연구

*문혜정 **김성경

아이엘피, 서울과학기술대학교

*hyejung.moon@gmail.com **gbible@naver.com

Study of Policy through Big data Analysis about Gambling News

*Moon, HyeJung **Kim, SungKyung

Institute of Lottery Policy, Seoul National University of Science & Technology

요 약

본 연구는 사행산업의 분야인 복권, 체육진흥투표권, 경마, 카지노에 대해 언론에서는 어떻게 다루어지고 있는지를 1990년부터 2015년까지의 뉴스데이터를 빅데이터 분석 방법 중 텍스트의 의미연결망 분석을 통해 밝혀보고자 하는 연구이다. 이 논문은 의미망 분석을 통해 기사의 빈도와 연결성을 프레이밍과 시민관심 정도로 재조명 하여 기사에 대한 언론보도자의 의도와 시민의 인식차이를 밝혔고, 이를 통해 정책적 특성과 개혁과제를 탐색하였다.

분석결과 복권의 경우 당첨번호, 당첨금, 조작의혹 등 당첨에 대한 부분이 주제인 '사회문제' 형태였으며, 체육진흥투표권의 경우에는 사업입찰, 불법사이트, 발매대상 등 주로 사업추진과 불법사이트에 대한 '의무정보' 종류였고, 경마의 경우 사업장, 홍보, 기사 등으로 사업홍보나 광고 관련 뉴스이었고, 마지막으로 카지노의 경우에는 불법, 도박장, 외국인 등 '주요정보'에 해당하는 논문이었다. 시대에 따라 1990년대에는 카지노, 2000년대에는 복권, 2010년대에는 경마에 대한 기사보도가 많아졌으며, 이에 대한 시민의 반응도 사업비리, 당첨, 시민운동 등의 차이가 있었다. 마지막으로 기사의 빈도와 연결성이 나타내는 프레이밍 정도와 시민의 관심은 '1. 홍보광고, 2. 의무정보, 3. 사회이슈, 4. 주요정보' 네 가지로 구분되었으며 이 중 사고, 비리 등 주요기사로 구분되는 사회문제가 주요 공공의제로 형성되는 것을 확인할 수 있었다.

주제어: 사행산업, 뉴스, 정책, 빅데이터, 분석

1. 서론

우리나라 사행산업의 매출은 사행산업통합감독위원회의 통계에 따르면 2014년 19조 8,933억원으로 전년 매출액 19조 6,726억원보다 1.1%(2,207억원) 증가한 수치로 2006년 이후 9년 연속 증가세에 있다 [1]. 이는 사행산업에 대한 관심이 높다 할 수 있다. 로또의 경우 발행 후 당초 예상보다 10배의 매출이 발행 하는 등 일명 로또 열풍을 일으키기도 하였다.

언론의 대표적인 매체인 뉴스를 살펴보다라도 전체적인 관심이 증가함을 할 수 있다. 미디어가온(www.mediagaon.or.kr)에서 제공하는 우리나라 전국 일간, 주간지 및 인터넷 뉴스, TV뉴스 등 총 61개의 언론사의 뉴스에 대해 1990년도부터 2015년도까지의 전체 뉴스에 대한 대표적인 사행산업인 로또, 스포츠토도, 경마, 카지노에 대한 뉴스 보도에 대한 비율을 나타내고 있다.

언론의 사행산업에 대한 관심이 증가 하고 있는 가운데 언론이 바라보는 사행산업은 어떠한 모습일까? 사행산업유형에 따라 어떠한 차이가 있을까? 이러한 맥락 속에서 위와 같은 질문을 해결해 보고자 사행산업 분야별 보도된 뉴스의 제목을 대상으로 텍스트 마이닝을 통한 단어의 빈도 분석과 텍스트 네트워크 분석을 통한 단어들 간의 네트워크 구조를 통해 사행산업 분야별 인식 대한 논의를 진행 하였다.

2. 선행연구고찰 및 이론적 배경

텍스트의 분석을 통한 인식이나 경향 등에 대한 연구는 빅데이터 분석과 더불어 언론학 분야에 점차 증가하는 추세이다. 소셜 네트워크 서비스(SNS)를 활용한 분석 사례로는 관광객들의 관광인식을 SNS 빅데이터를 통해 분석하는 연구[6]와 SNS의 텍스트에 대해 빅데이터를 활용한 국민의 통일에 대한 인식의 동향을 연구도 있었으며[5], '문학치료'에 대한 온라인상의 SNS자료에 대해 텍스트 마이닝과 오피니언스 마이닝 기법을 활용하여 문학치료와 관계된 주제의 내용을 분석한 연구도 있었다[9].

논문문헌을 대상으로 텍스트를 분석한 연구로는 최근 7년간의 문헌정보학 분야에 게재된 논문을 대상으로 빈도 분석과 네트워크 텍스트 분석을 한 연구가 있었으며[8], 신문기사를 통해 인식의 변화를 연구로는 관광기사를 분석하여 이를 통해 한국관광의 논제와 관광에 대한 인식에 연구가 있었다[3].

또한 신문기사의 텍스트를 대상으로 하여 의미연결망분석(semantic network analysis)을 통한 연구로는 텍스트 마이닝 기법을 활용하여 신문기사의 주제어 단순빈도 분석과 Clustering, Classification 결과를 분석한 연구[1]와 신문기사의 제목을 대상으로 빈도분석과 의미연결망분석을 통해 시기별 개인정보 유출의 프레임의

형태 변화를 연구한 사례가 있었으며[7], 빅데이터 관련 신문기사를 대상으로 의미연결망분석을 통해 새로운 기술과 현상인 빅데이터에 대한 언론은 어떻게 보도하고 있는지를 파악하고자 하는 연구가 있었다[10]. 이렇게 텍스트를 통한 의미 분석은 다양한 곳에서 이루어지고 있으며, 특히 신문기사를 통한 의미 분석은 주제어를 통해 수집된 텍스트 데이터의 의미연결망분석 통해 해당 주제에 대한 신문언론의 프레임링을 파악하고자 한다.

의미연결망분석(semantic network analysis)은 사회연결망분석(social network analysis)을 의사전달을 위한 메시지에 적용시킨 방법으로 사회연결망분석은 사회현상을 만들어내는 원동력을 각각의 행위자가 아니라 이들의 관계성에 주목하여 밝혀내는 방법이며, 의미연결망분석은 행위자 대신 메시지의 구성요소인 단어들의 구조적 관계에 주목 함으로서 이에 대한 의미를 파악하고 분석 한다[12].

3. 연구설계

사행산업 관련 뉴스의 빅데이터 분석을 통해 확인하고자 하는 주요 연구질문은 다음과 같다.

- 사행산업 관련 뉴스의 주요내용은 사행산업의 분야에 따라 어떠한 차이가 있는가?
- 사행산업 관련 뉴스는 주제에 따라 프레임링(빈도, 의도)과 네트워크(연결성, 중심성)의 특징은 어떠한가?
- 사행산업 관련 뉴스가 공공의제로 형성된 경우 대상 뉴스들의 프레임링과 네트워크의 특징은 어떠한가?

대중매체 특히 뉴스에 보도 되는 내용은 언론이 형성해내는 프레임링에 의해 많이 좌우된다. 그러나 뉴스에 대해 시민이 어떻게 반응하느냐에 따라 뉴스는 재보도 되고 심화되고 확산된다. 즉 뉴스는 초기에 소수 주요 언론사에 의해 프레임링 되나 이 뉴스가 견고해 지고 확산되는 것은 시민의 반응에 의해 차이가 생긴다는 것이다. 이 것은 사회연결망 이론에서 중심성과 연결성의 차이를 가진다. 이러한 전제를 위주로 사행산업 관련 뉴스를 통해 시민의 인식을 찾아보고자 한다.

분석 대상은 35년간 총 61개 언론사의 기사를 축적한 미디어가온(www.mediagaon.or.kr)의 뉴스 데이터베이스에서 사행산업 관련 주제어를 기초로 1990년도부터 2015년 까지 총 33,749,890건의 뉴스를 대상으로 하였다. 이중 우리나라 사행산업의 대표적인 분야를 기준으로 하여 검색 시 다른 해당 주제어에 가장 부합되는 데이터가 나오는 주제어를 선별하여 뉴스의 제목을 기준으로 복권(18,963건), 체육진흥투표권(2,277건), 경마(10,720건), 카지노(11,189)로 검색된 뉴스의 제목을 대상으로 하였다. 사행산업분야별로 검색된 기사의 건수를 년대별로 보면 1990년대의 경우 로또나 스포츠도트의 경우 기사가 거의 없거나 없었으며, 경마나 카지노의 경우에도 건수가 그리 많지는 않아 1990년대 이전의 뉴스는 분석대상에서 제외하였다.

분석절차는 1. 수집된 뉴스의 형태소 분석 처리 후, 2. 정제된 단어들을 시각적으로 표현 하기위해 각각의 기사별로 단어들의 동시출현 행렬을 구한 다음, 3. 이를 네트워크 데이터 시각화 툴인 NoadXL을 사용하여 연결정도의 중심성(degree centrality)을 도출 하였다. 5. 마지막으로 단어묶음의 군집을 확인하기 위해 활용 하여 데이터를 시각화

하여 의미망 분석을 수행하였다. 여기에서는 주제어는 모든 기사에 출연하는 인자 이므로 제외 하였고, 출현 빈도가 2개 이하의 빈도의 단어들의 경우 보다 정제된 형태의 데이터의 시각화를 위해 제외 하였다.

뉴스는 주요 소수의 언론사에 의해 먼저 보도된다. 그러나 이러한 뉴스가 견고해지고 확산되려면 시민이 뉴스에 대해 어떻게 반응하느냐에 따라 차이가 발생한다. 이는 다시 말해 특정 주제에 대해 뉴스가 얼마나 보도되었는지를 나타내는 빈도는 주로 언론사에 의해 결정된다. 반대로 뉴스가 다른 뉴스에 얼마나 연결되었는지를 나타내는 중심성은 뉴스가 처음으로 보도되고 나서 재인용되고 확산되었는지 정도를 나타낸다. 이는 기존의 뉴스가 확산되고 견고해진다는 의미는 시민이 뉴스를 보고 얼마나 반응했느냐의 정도를 나타내는 것이다. 이 연구에서는 이러한 뉴스의 보도건수와 연결성정도를 의도된 프레임링과 자연적으로 반응된 시민의식으로 다시 재해석해서 분석틀을 설계하였다. 이 연구의 분석틀은 <그림 1>과 같다.



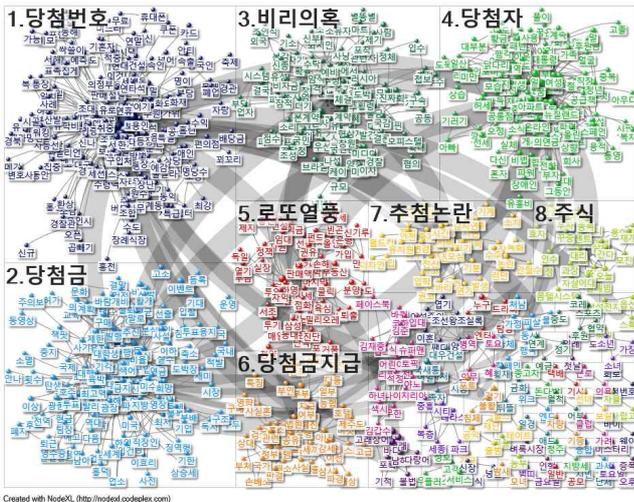
<그림 1. 분석틀>

뉴스의 보도정도와 확산정도의 상관성을 정리하면 '1. 홍보광고, 2. 의무정보, 3. 사회이슈, 4. 주요정보' 네 가지로 구분할 수 있다. 1. 홍보광고는 정보의 확산을 목적으로 언론사가 뉴스를 자주 배포하였으나 이에 대한 시민의 관심이 낮은 의도된 프레임링 정보일 가능성이 높다. 2. 소수정보는 언론사도 확산의지가 없고 시민의 관심도 낮은 소수인을 대상으로 한 의무적으로 보도되어야 하는 의무정보일 가능성이 높다. 3. 사회이슈는 언론사에서도 정보를 확산시킬 의사가 높고 이에 대한 시민의 반응도 높아 뉴스가 계속 확산되는 대중적인 정보일 가능성이 높다. 마지막으로 4. 주요정보는 언론사에서는 보도할 의도는 낮으나, 오히려 시민의 관심이 높아서 뉴스가 처음에 적게 보도되더라도 추후에 계속 확산되고 인용되는 경우가 많은 사회에서 중요한 이슈가 될 가능성이 높은 뉴스이다. 특히, 마지막 4. 주요정보는 사회에서 변화를 요구하는 주요 개혁과제가 될 가능성이 높다. 이러한 전제를 기준으로 뉴스의 의미망 분석을 통해 사행산업에 대한 시민의 인-식과 주요 정책이슈를 찾아보도록 하겠다.

4. 뉴스분석 결과

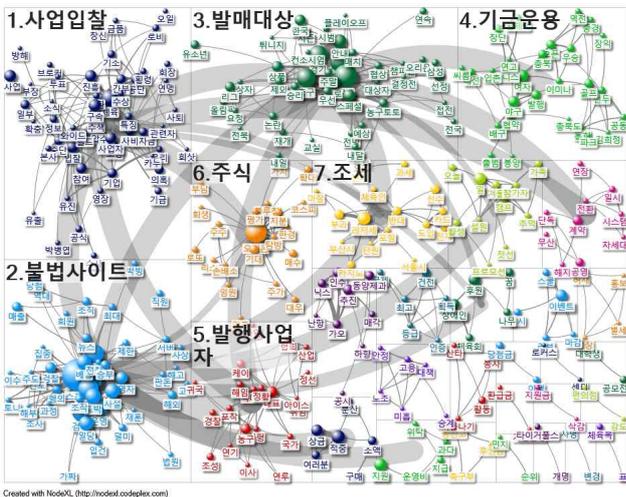
연구설계에서 설명한 바와 같이 주제별 빈도와 중심성은 뉴스의 보도정도와 시민의 반응정도를 나타낸다. 즉 주제어별로 빈도가 높다는 의미는 언론사에 의해 보도된 건수가 많다는 것이다. 이를 사행산업분야별로 의미망 분석을 수행하여 뉴스 간 연계되어 함께 보도된 언론사의 의도를 표현하는 프레임링을 네트워크 분석에서 그룹으로 표현하고, 뉴스 간에 어떻게 확산되었는지를 네트워크의 연결성으로 표현

하여 그룹 내에서 중심에 있는 뉴스의 주제를 연결중심성이 높은 것으로 도식화하였다.



<그림 2. 사회이슈 형태의 복권 관련 뉴스>

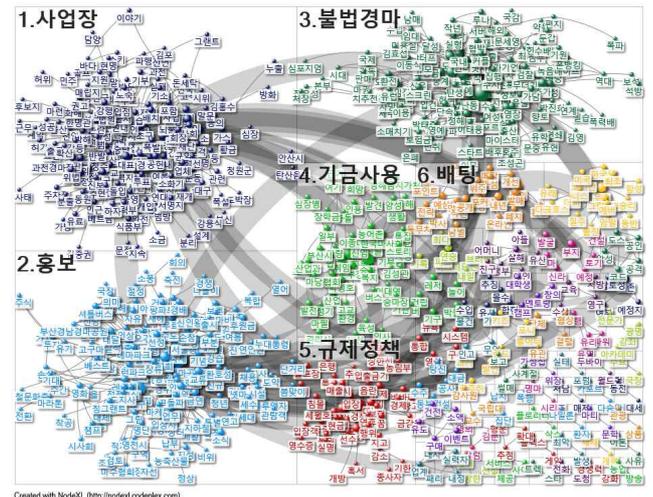
복권 분야의 뉴스는 보도건수도 많고 이에 이후 뉴스가 확산된 정도도 높다. 보도가 많이 된 뉴스를 주제별(건수)을 보면 '당첨(5,039), 번호(3,851), 보너스(1,365)' 등 당첨결과에 관한 보도이며, 뉴스간에 확산성이 높은 주제의 뉴스(연결중심성)는 '대박(28,512), 명당(6,871), 행운(3,637)' 등 희망적인 내용이 주를 이루어 가장 보도건수가 많은 분야로 확인되었다. 관련 정보를 확산시키고자 하는 언론사의 의도가 높았으며, 뉴스에 대한 시민의 관심도 컸다는 것을 암시한다. 이러한 '사회이슈'형 뉴스가 '공공의제'로 형성된 것은 2003년 로또복권발매 이후 시민의 의도가 높아지자 대한민국 최초로 '복권및복권기금법(2004)'을 제정하고, '복권위원회'를 설립한 것이다. 이로써 2004년 복권만을 위한 법제도와 기구가 생성되었다.



<그림 3. 의무정보 형태의 토토 관련 뉴스>

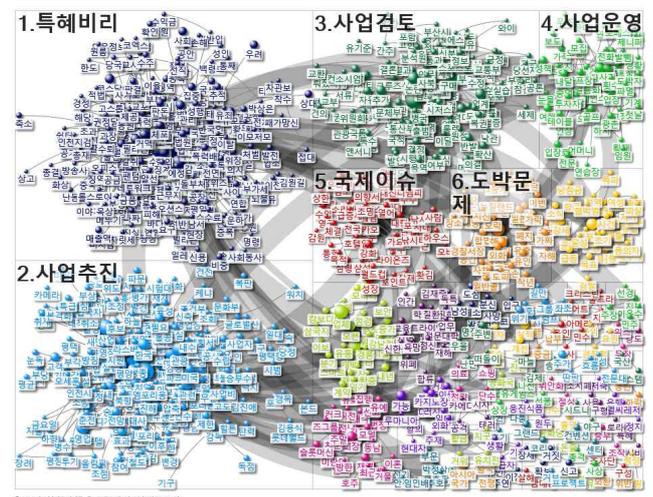
스포츠도도의 경우 보도된 뉴스의 건수는 다른 사행산업 분야에 비해 매우 적다. 보도가 많이 된 뉴스(건수)는 '불법(296), 발매(178),

오리온(152)' 등 사업비리에 대한 내용이고, 확산된 뉴스(연결성)은 '축구(5,374), 적발(1,363), 구속(48), 여자(47)' 등 스포츠도도 결과의 조 작관련 뉴스로 사행산업 중 보도건수가 가장 적은 분야였다. 이는 애초에 관련된 뉴스의 확산을 막으려는 의도를 다분히 확인할 수 있다. 반면 적은 뉴스보도인 만큼 확산된 정도는 다른 뉴스보다는 상대적으로 작다. 이러한 뉴스는 반드시 시민에게 공개해야 하는 뉴스를 최소화하고자 프레임링 하는 '의무정보'의 형태로 이해할 수 있다.



<그림 4. 홍보광고 형태의 경마 관련 뉴스>

경마분야에서 뉴스는 보도건수는 높으나 상대적으로 확산된 정도는 낮은편이다. 주된 보도(건수)는 경마장(2087), 일(1198), 경주(1035), 부산(1009)' 등 경마장 시설 관련한 보도가 주류였으며, 확산성이 높은 뉴스(연결중심성)은 '기수(23,233), 말(9,755), 조교사(4,975)' 등 전문성과 관련된 뉴스였다. 이는 주로 홍보나 광고 등 시민에게 특정 정보를 제공하고자 하는 언론사의 의도로 확인할 수 있다. 그러나 이러한 홍보성 정보에 대한 시민의 반응은 그리 높지 않았던 것으로 확인된다.



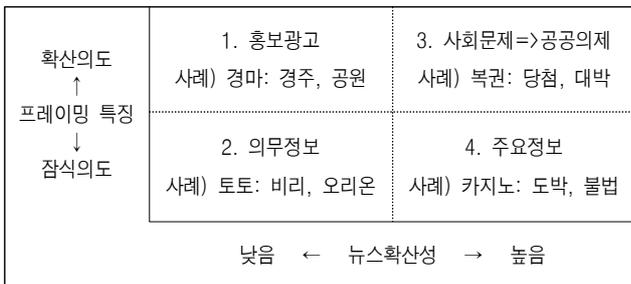
<그림 5. 주요정보 형태의 카지노 관련 뉴스>

카지노 관련 뉴스는 보도건수는 상대적으로 낮으나, 특정 뉴스의

확산정도는 높은 것으로 확인되었다. 이 분야에서 보도건수가 높은 뉴스(건수)는 '도박장(1,094), 강원랜드(700), 외국인(638)' 순서였으며, 중심성이 높은 뉴스는(연결중심성)은 '마카오(28,935), 내국인 출입(41,116), 내국인허용(15,075)' 관련 뉴스였다. 즉 언론사는 해당 뉴스의 정보를 잠식시키려 하나 특정 이슈에 대해서 시민들의 관심이 높아 해당 뉴스가 사회문제화 된 것으로 이해된다.

5. 결론 및 시사점

사행산업 관련 뉴스를 보도건수와 언론사 의도 기반의 프레임 특징과, 네트워크 연결성 기반의 시민 관심정도로 구분하면 <그림6>과 같다.



표기: 1.사행산업분야, 2. 기사형태

<그림 6. 사행 분야별 뉴스 프레임링과 연결성>

사행산업 관련 기사 분석 결과를 나누면 '1. 홍보광고, 2. 의무정보, 3. 사회문제, 4. 주요정보' 네 가지이다. 이중 공공의제화 된 뉴스의 분야는 사회문제 형태이다. 즉 언론사가 정보를 확산하려는 의도도 확실하고, 이에 대한 시민의 반응이 높아야 공공의제로 형성되는 것이다. 이는 2006년 연합복권에 대한 의제설정 당시, 사회문제에 대한 정보의 개입의도가 높고 시민의 관심이 높을 경우 의제설정이 공고(Consolidation)히 된 복권정책의 사례와도 유사한 양상을 보인다[2].

본 논문은 지난 30년간 사행산업 관련한 뉴스를 근거로 사행산업에 대한 언론보도프레임링과 시민인식의 차이와 변화를 분석해보았다. 뿐만 아니라 의미망 분석결과를 토대로 기사의 빈도와 연결성을 프레임링과 시민관심 정도로 재조명 하였고, 이를 통해 정책적 특성과 개혁과제를 찾아볼 수 있었다. 기존의 뉴스분석에서 의미망 분석을 통한 인식변화에 대한 연구는 많았다. 그러나 네트워크 빈도와 연결성을 정보의 중요성과 시민의 반응으로 다시 해석하여 정책의제의 특성으로 해석한 시도는 이 논문이 처음이다. 향후엔 이 논문을 주제로 실재 정책의제로 선정된 사례와 정책문제로 선정되어 해결된 사회문제가 있는지 심도 있게 연구해 보고자 한다.

참고문헌

[1] 감미아·송민 (2012) “텍스트 마이닝을 활용한 신문사에 따른 내용 및 논조 차이점 분석”, 지능정보연구, 18(3): 53-77.

[2] 문혜정. (2012a). 한국 복권정책의 의제설정에 관한 연구. 「한국사회와 행정연구」 23(1): 287-316.

[3] 민경익·이태숙 (2012) “미디어분석을 통한 한국관광 인식 변화에 관한 연구 - 신문기사를 중심으로”, 관광학연구, 36(10): 81-104.

[4] 사행산업통합감독위원회 (2014) “2014년 사행산업 관련 통계”, 국무총리실, 11-1371045-000037-10.

[5] 송태민 (2015) “소셜 빅데이터를 활용한 국민 통일인식 동향 분석”, Issue & Focus, 269: 1-8.

[6] 오익근·이태숙 (2015) “빅데이터 분석을 통한 한국관광 인식에 관한 연구”, 관광학연구, 39(10): 107.

[7] 정서화·조현석 (2014) “사회연결망 분석을 활용한 개인정보 유출 프레임 변화에 관한 연구 : 1984년-2014년을 중심으로”, 디지털융복합연구, 12(5): .57-68.

[8] 조재인 (2011) “네트워크 텍스트 분석을 통한 문헌정보학 최근 연구 경향 분석”, 정보관리학회지, 28(4): 65-83.

[9] 최경호·박정혜 (2015) “빅데이터 분석을 통한 문학치료의 대중적 인지도 분석 - 국문학과 통계학의 융합적 측면”, 디지털 융복합연구, 13(4): 395-404.

[10] 최윤정·권상희 (2014) “‘빅데이터’ 관련 신문기사의 의미 연결망 분석”, 사이버커뮤니케이션 학보, 31(1): 241-286.

[11] Entman, R.M.(1993), “Framing: Toward clarification of a fractured paradigm”, Journal of Communication, 43: 51~58.

[12] Wang, W. & Rada, R.(1998), “Structured hypertext with domain semantics”, ACM Trans. Inform. Syst., 16: 372~412.

[13] Wasserman, S. & Faust, K.(1994), “Social network analysis: methods and applications”, Cambridge University Press.