

데이터사이언스 활용의 현재와 미래: 한국방송학회 발간 학보를 중심으로

한양대학교 언론정보대학
정보사회미디어학과
이창준

1. 도입

2. 문제 제기

3. 연구 방법

4. 결과

5. 토의

Computational Social Science

≈ Social Science with **Data Science** tools

Structured data



Unstructured data

In tidy data:

- each variable forms a column
- each observation forms a row
- each cell is a single measurement

each column a variable

id	name	color
1	floof	gray
2	max	black
3	cat	orange
4	donut	gray
5	merlin	black
6	panda	calico

each row an observation

Wickham, H. (2014). Tidy Data. Journal of Statistical Software 59 (10). DOI: 10.18637/jss.v059.i10

- Text (from Documents, Web, SNS)
- Image (from Documents, Web, SNS)
- Relational DB (Bigdata DB, open API etc.)

Computational Social Science

≈ Social Science with **Data Science** tools

- Import & Crawling *Technique*
- Pre-processing *Technique*
- Visualization *Technique*
- **Analysis & machine learning** *Science*
 - Modelling
 - Parameter tuning & Validation
- Prediction *Technique*

Science

사회과학에서 데이터 사이언스 활용 → Computational Social Science (계산 사회과학)

아래 네 가지로 정의되는 사회적 탐구 (Social inquiry)에 대한 접근 방식

1. (항상 그런 것은 아니지만) 테라바이트 또는 페타바이트 단위로 측정되는 크고 복잡한 데이터 (large and complex) 사용 → 크거나 복잡하거나 둘 중 하나만 만족하기도 함.
2. "자연적으로 발생하는 (*naturally occurring*)" 소셜 및 디지털 미디어 소스 및 전자 데이터베이스의 사용
3. 이러한 데이터로부터 패턴과 추론을 생성하기 위한 계산 또는 알고리즘 솔루션의 사용 (ML)
4. 다양한 영역에서 사회 이론에 적용할 수 있음 (i.g. 대중의 의견 연구에서 공중 보건 이슈, 정치 사건 조사에서 사회 운동에 이르기까지)

→ 위 네 가지 요소 중에서 “자연 발생 데이터 사용, *naturally occurring*” 부분이 특별히 중요함: 전통적인 설문과 실험 방법과 대조적이며 이들과 상호 보완이 가능하기 때문

사회과학에서 데이터 사이언스 활용 → Computational Social Science (계산 사회과학)

전통적 방법의 설문(Survey)과 실험(Experiment)

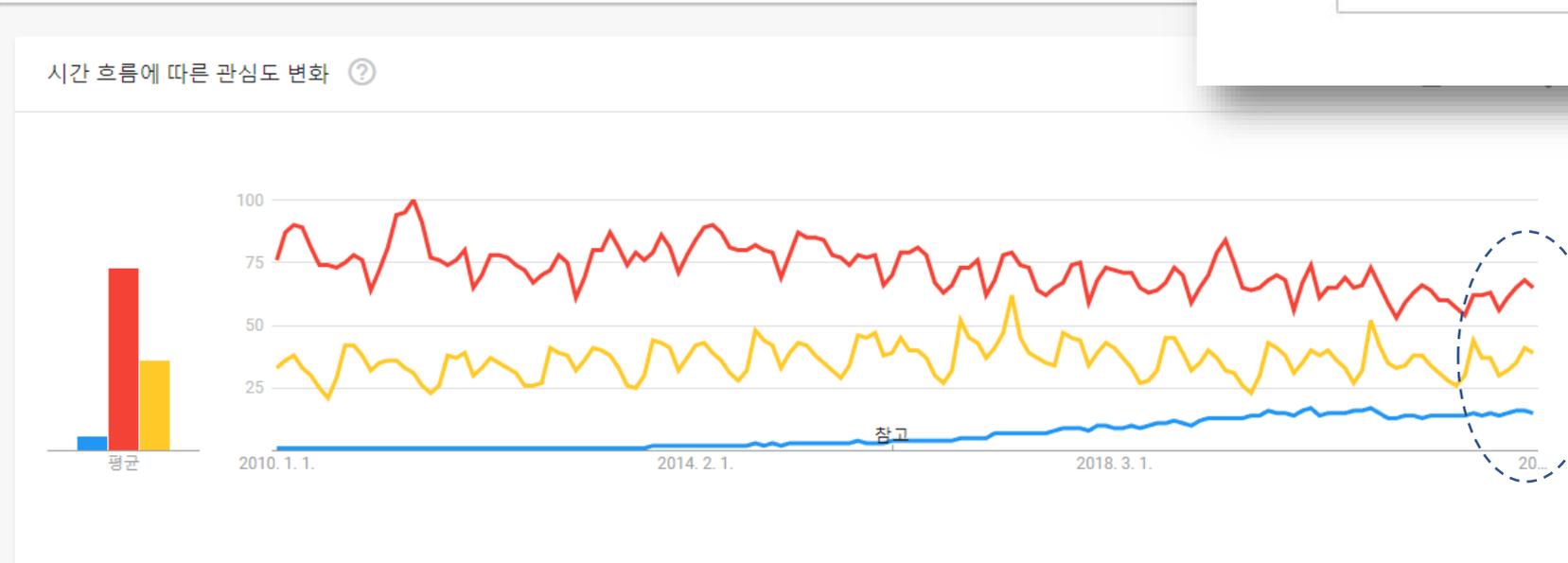
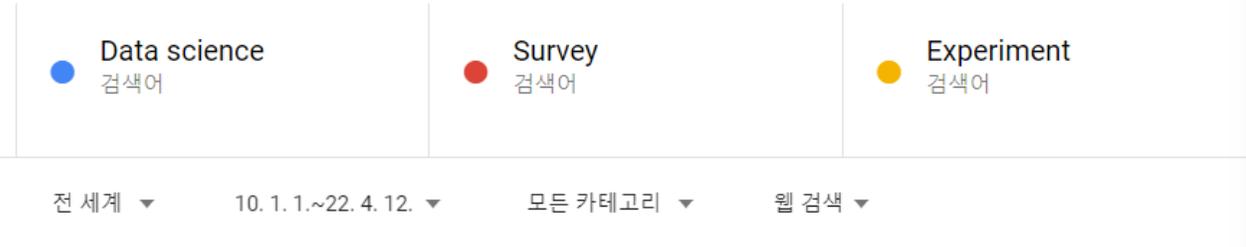
1. 설문과 실험에서는 어쩔 수 없이(by their nature) ‘연구자’가 사회적 프로세스 (또는 상황)에 개입
→ 실험자 효과(Experimenter effect) 발생: 대조군과 실험군에 (비자발적으로) 주관적 편이가 존재
2. 또한 자기-경험에 의존한 진술 편이 (self-reported bias) 내재
(예를 들어 설문자가 응답시 good-quality behavior 에는 상향 진술을, 바람직하지 않은 행동에는 하향 진술을 하는 경향이 존재)

계산사회과학은 위의 단점을 보완할 수 있는 ‘대안(보완)적’ 역할을 담당할 수 있음.

→ 하지만, 일반화의 어려움(이론 부재), 데이터 취득과 저장에 있어 윤리적 기술적 문제, 전통적인 데이터 베이스 방법으로 처리가 불가능한 비정형 데이터 처리의 어려움 등은 여전히 한계점

Google Trend [2010 ~ 2022]

- 데이터사이언스(Data science): 6
- 설문(Survey): 73
- 실험(Experiment): 36



감정도 전염되는 것일까?

Experimental evidence of massive-scale emotional contagion through social networks

Adam D. I. Kramer^{a,1}, Jamie E. Guillory^{b,2}, and Jeffrey T. Hancock^{b,c}

^aCore Data Science Team, Facebook, Inc., Menlo Park, CA 94025; and Departments of ^bCommunication and ^cInformation Science, Cornell University, Ithaca, NY 14853

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved March 25, 2014 (received for review October 23, 2013)

Emotional states can be transferred to others via emotional contagion, leading people to experience the same emotions without their awareness. Emotional contagion is well established in laboratory experiments, with people transferring positive and negative emotions to others. Data from a large real-world social network, collected over a 20-y period suggests that longer-lasting moods (e.g., depression, happiness) can be transferred through networks [Fowler JH, Christakis NA (2008) *BMJ* 337:a2338], although the results are controversial. In an experiment with people who use Facebook, we test whether emotional contagion occurs outside of in-person interaction between individuals by reducing the amount of emotional content in the News Feed. When positive expressions were reduced, people produced fewer positive posts and more negative posts; when negative expressions were reduced, the opposite pattern occurred. These results indicate that emotions expressed by others on Facebook influence our own emotions, constituting experimental evidence for massive-scale contagion via social networks. This work also suggests that, in contrast to prevailing assumptions, in-person interaction and non-verbal cues are not strictly necessary for emotional contagion, and that the observation of others' positive experiences constitutes a positive experience for people.

demonstrated that (i) emotional contagion occurs via text-based computer-mediated communication (7); (ii) contagion of psychological and physiological qualities has been suggested based on correlational data for social networks generally (7, 8); and (iii) people's emotional expressions on Facebook predict friends' emotional expressions, even days later (7) (although some shared experiences may in fact last several days). To date, however, there is no experimental evidence that emotions or moods are contagious in the absence of direct interaction between experimenter and target.

On Facebook, people frequently express emotions, which are later seen by their friends via Facebook's "News Feed" product (8). Because people's friends frequently produce much more content than one person can view, the News Feed filters posts, stories, and activities undertaken by friends. News Feed is the primary manner by which people see content that friends share. Which content is shown or omitted in the News Feed is determined via a ranking algorithm that Facebook continually develops and tests in the interest of showing viewers the content they will find most relevant and engaging. One such test is reported in this study: A test of whether posts with emotional content are more engaging.

The experiment manipulated the extent to which people (N = 689,003) were exposed to emotional expressions in their News Feed. This tested whether exposure to emotions led to

computer-mediated communication | social media | big data

<Study Design>

- Facebook 에서 긍정적/부정적 post 노출 → 향후 해당 ID 의 포스팅의 언어 분석을 통해 감정의 전이를 측정
- 대략 700,000명의 개인(확인된 유저 ID)을 대상으로 두 개의 실험을 한 주 가량 병렬로 진행
- (1) In one experiment, 대조군에 비해 긍정적 감정의 콘텐츠를 10~90% 감소시킴
- (2) In the other experiment, 부정적 감정의 콘텐츠를 10~90% 감소시킴

<Study Result>

- 긍정적 콘텐츠 노출 감소
→ 긍정 단어 사용 0.1% 감소 / 부정 단어 0.04% 증가
- 부정적 콘텐츠 노출 감소
→ 부정 단어 사용 0.07% 감소 / 긍정 단어 0.06% 증가

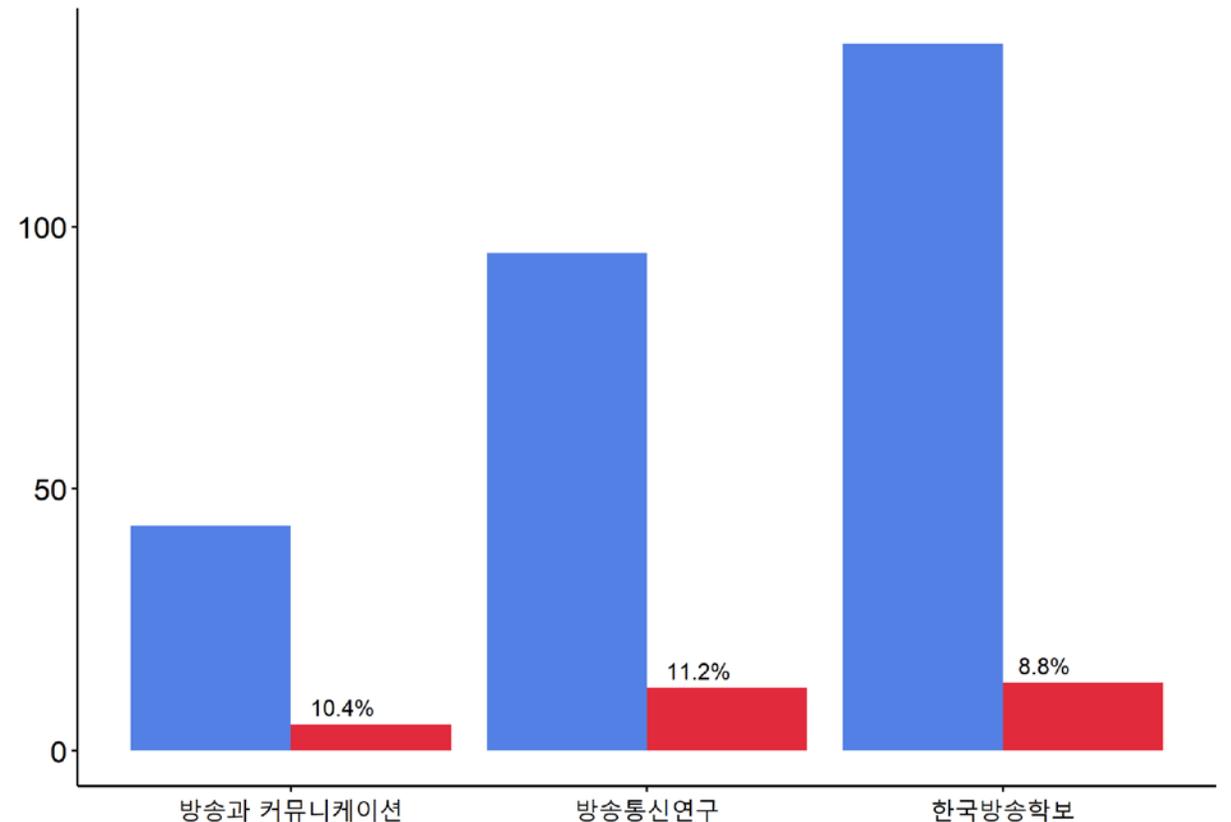
- 해외에서는 이러한 계산사회과학이 일찍이 사회과학의 여러 분과에 자리 잡아가고 있음.
 - 특히 커뮤니케이션 분과는 분과 내에 데이터사이언스 트랙이 있거나 커뮤니케이션 분과 자체가 정보 대학(Information school)과 같은 우산 아래에 위치해 있기도 하는 등
 - 경영학과 내의 정보시스템(Information System)과 더불어 사회과학에서 데이터사이언스 접목의 선두 주자 역할을 담당하고 있음.
- 해외 학술지에서도 2010년대 초반에 빅데이터 또는 데이터사이언스 활용에 대한 특별 이슈(Special issue)를 배치하는 등 활발한 융합 시대를 열어왔음.
 - 그 이후 현재까지 계산사회과학은 정보통신기술의 발달과 함께 데이터사이언스로의 융합은 한철 유행이 아닌 거스를 수 없는 대세로 여겨지고 있음.
 - 다만 그 안에서 키워드만 텍스트 데이터 분석에서 VR/AR, 영상과 사운드 데이터 분석, 또는 메타버스 등으로 바뀌고 있음.

- 이 시점에서 우리는 국내 커뮤니케이션 연구 분야 특별히 방송학이라는 학문 분과 내에서 데이터사이언스의 활용의 현재와 미래에 대한 토론이 필요
- 이를 위해 방송학에서 데이터사이언스는 어떻게 융합되어 어떤 방향으로 진화하고 있는지에 대한 연구가 필요.
- 따라서 본 연구는 국내에 적극적으로 계산사회과학이 도입되기 시작한 2010년대 중후반부터 현재까지 방송학에 데이터사이언스가 어떻게 활용되고 있는지를 분석해 보고자 함.

- 데이터: 한국방송학회 발간 학보인 한국방송학회지, 방송통신연구, 방송과커뮤니케이션의 2017년부터 2022년까지 5개년의 연구 (총 연구 논문 수: 303)

저널	논문 수	DS (%)
한국방송학보	148	8.8
방송통신연구	107	11.2
방송과커뮤니케이션	48	10.4

Data-science as a tool
at each media-related journal



4. 연구 결과

- 연구 방법론 분포(모든 논문)

사례연구

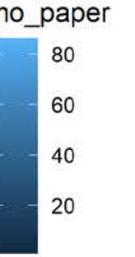
설문형태, 심층인터뷰	딤러닝, 머신러닝	실험설계	연관어분석, 감정분석			
		설문형태, 사회연결망분석		초점집단인터뷰		
사례연구, 심층인터뷰		설문형태, 사례연구		의미연결망분석, 키워드추출	텍스트분석, 관계도분석	
	다양성분석, 생존분석	설문형태, 내용분석			의미연결망분석, QAP분석	
	내용분석, 브레인스토밍	설문형태, G-K오형	심층인터뷰, 내용분석		심층인터뷰, 설문형태	
			메타분석		사회연결망분석	설문형태, CULDIFF분석
	관찰방법	내용분석, 심층인터뷰	시나리오분석	심층인터뷰, 관찰방법		

심층인터뷰

설문형태, 실험설계

설문형태

내용분석



• 데이터사이언스를 활용한 논문 샘플의 연구 방법론 분포



4. 연구 결과

- 데이터사이언스를 활용한 논문 샘플의 연구 방법론 연도별 변화 [2018:2022]

2018

연관어분석, 감성분석

토픽모델링LDA, 주성분분석, 토픽모델링알고리즘, 토픽맵

사회연결망분석

2019

다양성분석, 생존분석

텍스트분석, 의미연결망분석

로지스틱회귀분석, 머신러닝(로지스틱리그래션모델, 의사결정나무모델)

내용분석

다차원척도법, K평균군집분석, 토픽모델링, LDA

no_paper

1

4. 연구 결과

- 데이터사이언스를 활용한 논문 샘플의 연구 방법론 연도별 변화 [2018:2022]

2020



4. 연구 결과

- 데이터사이언스를 활용한 논문 샘플의 연구 방법론 연도별 변화 [2018:2022]

2021

설문형태, CULDIFF 방정식	텍스트분석, 관계도분석	패널분석(고정효과모형)
딥러닝, 머신러닝	텍스트마이닝, 형태소분석, 텍스트분석	토픽모델링LDA, 상호작용연결망분석, 텍스트연결망분석
감성지수분석, 오피니언마이닝, 사전기반분석	심층인터뷰, 텍스트데이터마이닝, 워드클라우드	의미연결망분석, QAP 분석

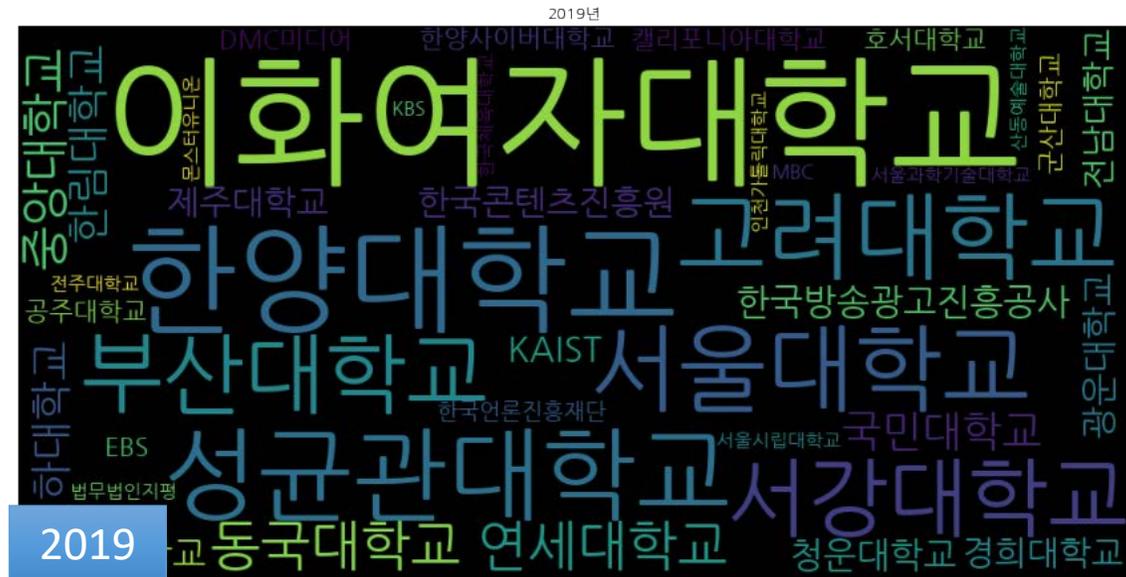
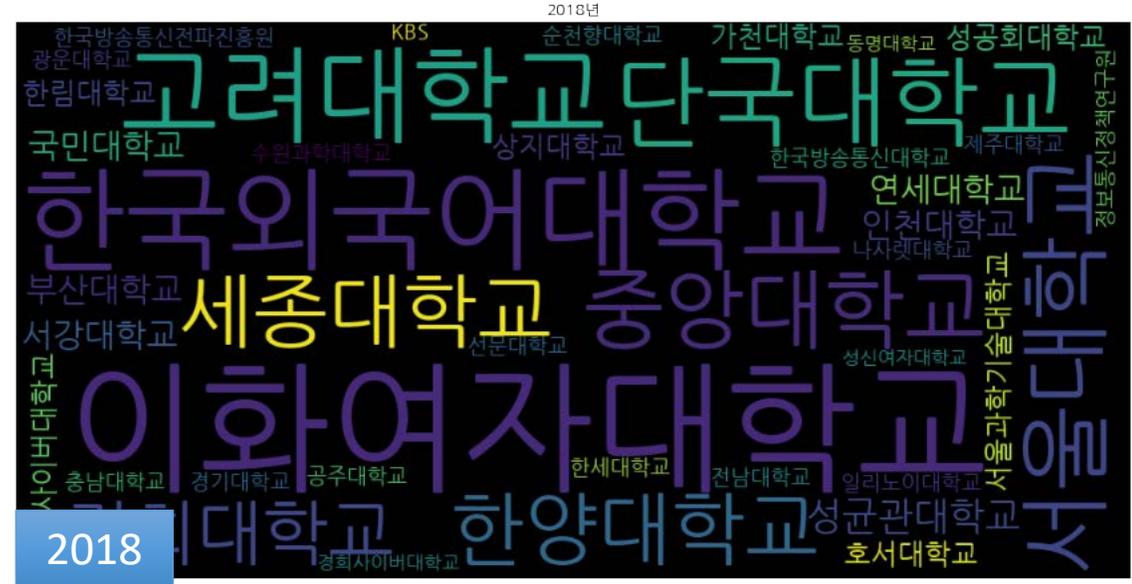
no_paper



1

4. 연구 결과

• 데이터사이언스 논문 투고 기관의 변화

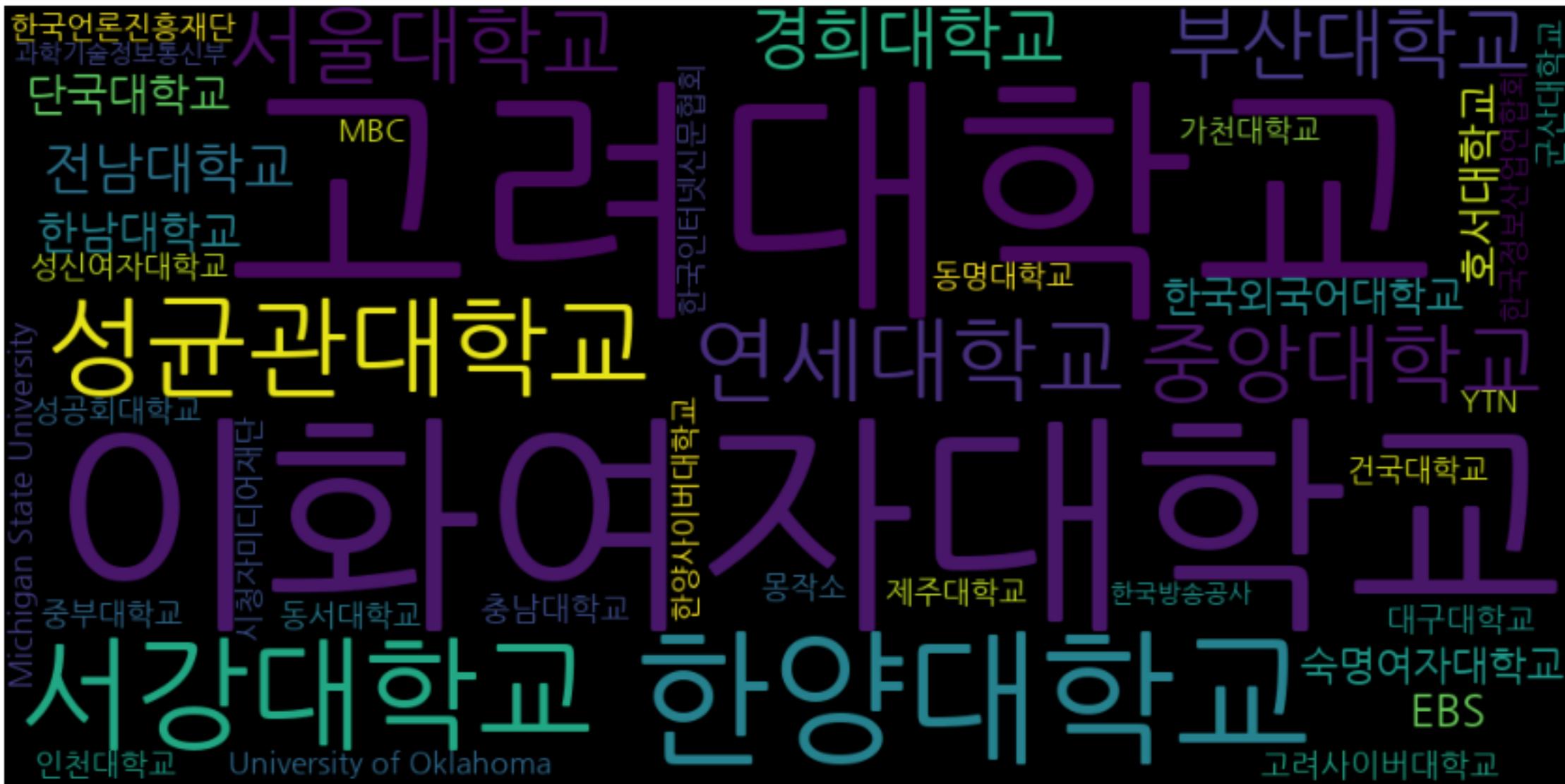


4. 연구 결과

- 데이터사이언스 논문 투고 기관의 변화

2021

2021년



- 방송학에서 데이터사이언스 방법론 활용 논문
 - 2015년부터 점점 증가하여, 약 10% 정도 차지할 정도가 되었음.
- 연구 방법론의 변화
 - 모든 논문에서 연구 방법론의 분포는 사례연구와 설문형태가 가장 많았고, 심층인터뷰, 실험설계, 내용분석이 뒤를 이었다.
 - 데이터사이언스를 활용한 연구로 한정했을 때는 내용분석이 가장 많았지만 감성분석, 다양성분석, 토픽모델링, 사회연결망분석 등 다양하게 분포되어 있었음
 - 2018년에서 20년까지는 텍스트 중심의 분석이 주를 이루었으나 2021년에 들어 딥러닝과 머신러닝을 활용한 예측 모형, 패널분석 등도 나타남
- 투고 기관의 변화
 - 논문 투고 기관으로는 2017년에는 고려대, 강원대, 한림대, 서울대학교가 주도하였고 2018년부터 이화여자대학교, 단국대학교, 한국외국어대학교가 씬에 활력을 주었음.
 - 2019년에는 여기에 한양대학교, 성균관대학교, 부산대학교가 새로이 강세를 보였고, 2020년에는 한양대와 중앙대, 서강대가 씬을 이끌었음.
- 키워드 변화
 - 방송통신, 인공지능 → 인터넷 개인방송, 사회적 시청, 광고노출효과 →
 - 소셜미디어, 이용동기, OTT, 종편, 고정관념 → 미디어, 유튜브, 공영방송, 다양성, 소수자 →
 - 넷플릭스, 유튜브, OTT, 선정성, 공정성, 갈등

- 현재까지 방송/언론학에서 데이터사이언스의 역할
 - 비정형 텍스트 데이터를 다루는 여러가지 시도에 불과함
 - 딥러닝 등 예측 모형을 적용하려는 적은 시도들이 있었지만 여전히 사회과학의 철학과는 다른 이유로 배척
- 미래 방송학에서 계산사회과학 발전을 위한 제언
 - 계산사회과학에 대한 인식 변화/개선을 위한 연구반 모임 활성화
 - 전통적인 틀에서 벗어난 특별호 편찬 빈도를 늘리는 노력 필요
 - 방송학과 다른 학문(e.g. 컴퓨터공학, 산업공학, 응용물리학, AI융합학 등)과의 협업에 인센티브 부여
 - 국내 학술지 리뷰어 POOL 에 계산사회과학 편집위원 다수 등용

- Shah, D. V., Cappella, J. N., & Neuman, W. R. (2015). Big data, digital media, and computational social science: Possibilities and perils. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 6-13.
- Chen, N. C., Drouhard, M., Kocielnik, R., Suh, J., & Aragon, C. R. (2018). Using machine learning to support qualitative coding in social science: Shifting the focus to ambiguity. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 8(2), 1-20.
- Grimmer, J., Roberts, M. E., & Stewart, B. M. (2021). Machine learning for social science: An agnostic approach. *Annual Review of Political Science*, 24, 395-419.
- Agarwal, R., & Dhar, V. (2014). Big data, data science, and analytics: The opportunity and challenge for IS research. *Information systems research*, 25(3), 443-448.
- Hindman, M. (2015). Building better models: Prediction, replication, and machine learning in the social sciences. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 48-62.
- Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788-8790.