

소셜네트워크서비스 빅데이터 분석을 위한 연구문제 설정과 통계적 제 문제-융합적 관점*

박한우*, 최경호**

영남대학교 언론정보학과*, 전주대학교 보건통계연구소**

Doing social big data analytics: A reflection on research question, data format, and statistical test-Convergent aspects*

Han-Woo Park*, Kyoung-ho Choi**

Department of Media & Communication, YeungNam University*

Research Institute of Health Statistics, Jeonju University**

요 약 타당한 연구 수행을 위해서는 방법론이 중요하다. 소셜네트워크서비스로부터 수집되는 데이터를 대상으로 하는 소셜 빅데이터 연구는 최근 들어 새롭게 부각되는 연구이지만 아직 이에 합당한 연구방법이 충분하지 않은 실정이다. 이에 본 연구에서는 소셜 빅데이터 분석에 합당한 연구방법론 개발에 앞서, 연구문제의 설정에 대하여 체계적으로 정리하고 질문의 기본 유형을 제시하고자 한다. 그리고 제시되는 6가지 기본 유형에 따른 데이터 형태를 살펴보고자 한다. 나아가 SNS로부터 수집되는 빅데이터 분석과 관련된 통계적인 제 문제에 대해서도 고찰해 보도록 하겠다. 본 연구의 결과는 향후 관련 연구자들이 데이터 유형에 맞는 올바른 연구문제를 수립하고 분석함으로써 타당한 정보를 도출하는데 도움이 될 것으로 사료된다.

주제어 : 소셜네트워크서비스, 빅데이터, 연구문제, QAP.

Abstract Research question and method play important roles in conducting a research in a scientifically valid way. In today's digitalized research environment, social network service (SNS) has rapidly become a new source of big data. While this shift provides new challenges for researchers in Korea, there is little scholarly discussion of how research questions can be framed and what statistical methods can be applied. This article suggests some basic but primary types of example questions for researchers employing social big data analytics. Further, we illustrate the interface of the intended data set specifically for SNS-mediated communication and information exchange behaviors. Lastly, a statistical test known as proper method for social big data is introduced.

Key Words : Social network service, big data, research question, QAP.

* The earlier version of this manuscript was presented by the first author in the Spring Seminar jointly organized by Korea Association for Data Science and World Association for Triple Helix & Future Strategy Studies (WATEF). It has been substantially revised in a collaboration with the corresponding author for journal publication.

Received 25 October 2016, Revised 30 November 2016

Accepted 20 December 2016, Published 28 December 2016

Corresponding author: Kyoung-ho Choi (Jeonju University)

E-mail: ckh414@jju.ac.kr

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

최근 스마트폰, 스마트TV, RFID, 센서 등의 급속한 보급과 모바일 인터넷과 소셜미디어의 확산으로 데이터량이 기하급수적으로 증가하고 데이터의 생산·유통·소비 체계에 큰 변화가 일어나면서 데이터가 경제적 자산이 될 수 있는 빅데이터 시대를 맞이하게 되었다[1]. 위키피디아에 따르면 빅데이터란, ‘기존 데이터베이스 관리 도구로 데이터를 수집·저장·관리·분석하는 역량을 넘어서는 대량의 정형 또는 비정형 데이터 세트 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술’이라고 정의되고 있다. 이와 같은 빅데이터 시대가 출현한 가장 큰 배경에는 디지털화의 진전 즉 디지털 기술의 발전과 기기의 보급을 들 수 있다. 새로이 부각되고 있는 디지털 기술의 발전에는 두 가지가 있는데, 그 중 하나가 바로 스마트폰과 SNS(social network service)의 확산이다[2]. 이에 따라 SNS에 기인한 소셜 빅데이터(social big data)를 이용하여 정보를 도출하고 의미 있는 결과를 분석하는 많은 연구들이 다양한 영역에서 발표되고 있다. 그 예로 SNS를 통한 참여형 재난관리체계의 구축에 대하여 연구한 [3], 트위터상의 의견들을 수집하여 오피니언 마이닝을 통해 소비자들이 소셜미디어에 표출한 반응이 기업이미지의 선행지표가 될 수 있음을 살펴 본 [4] 그리고 SNS 빅데이터를 이용하여 영화 마케팅 전략을 도출한 [5] 등을 들 수 있다.

한편 연구(research)란 증거가 없는 상식을 체계적·구체적·논리적인 방법으로 증거를 확인하여 이론을 정립해 주는 작업으로 [6], 모든 연구에는 합당한 연구방법이 있어야 한다. 즉 논문을 쓰거나 연구를 하기 위해서는 어떤 입장에 서야 한다. 동시에 타당한 논리를 구사해야 하며 이를 뒷받침하는 자료를 제시하고 분석하고 해석함으로써 어떤 주장을 펼 수 있게 되는데, 이것은 모두 방법론의 영역으로서 논리적 주장이나 연구가 그에 맞는 방법에 따르지 않으면 그 연구는 정상이 아니다. 이렇듯 연구에 있어서 방법론은 그만큼 중요한 필요충분조건이다[7]. 그런데 소셜 빅데이터를 활용한 연구는 최근 들어 새롭게 부각되는 연구로 아직 이에 합당한 연구방법이 충분히 논의되지 않은 실정이다[8]. 특히 SNS로부터 수집되는 데이터는 특성상 매트릭스 형태로 구성되는 바, 기존의 통계분석에서의 데이터와는 입력형식부터 상당

한 차이가 있다. 이에 본 연구에서는 소셜 빅데이터 분석에 합당한 연구방법론 개발에 앞서, 연구문제의 설정에 대하여 체계적으로 정리하고 질문의 기본 유형에 따른 데이터 형태를 살펴보고 나아가 SNS로부터 수집되는 빅데이터 분석과 관련된 통계적인 제 문제에 대해서도 융합(convergence)적인 관점에서 다루고자 한다.

2. 소셜 빅데이터 분석을 위한 연구문제의 설정

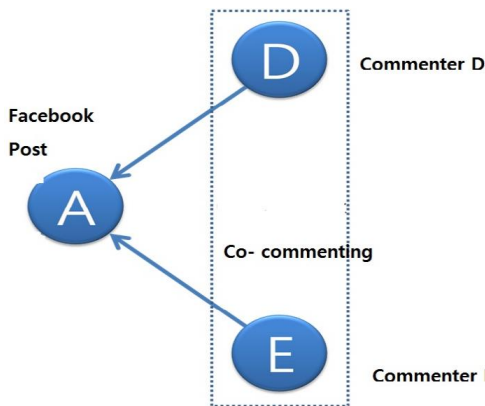
소셜 빅데이터 분석은 먼저 연구의 목적에 맞는 연구문제를 설정하는 것으로부터 시작된다. SNS는 사람들 간 정보와 정서 교류를 매개하고 촉진하는 채널이다. 이에 일반적으로 소셜 빅데이터 분석에 적합한 질문의 기본유형은 다음과 같다[9]. 첫째, 사람들이 무엇을 얘기하는가? (①) 둘째, 누가 무엇을 보고 주목하는가? (②) 셋째, 누가 누구에게 대답하는가? (③) 넷째, 콘텐츠가 얼마나 지속되는가? (④) 다섯째, 무엇이 무엇에 연결되어 있는가? (⑤) 여섯째, 누가 누구와 관계를 맺고 있는가? (⑥) 소셜 빅데이터 분석을 위한 연구문제는, 기본적으로 이와 같은 6가지의 질문유형을 토대로 작성되지만, 다음의 몇 가지 예에서 보듯이 상황에 따라 자유롭게 변형될 수 있다. 먼저 세계인터넷전문가총회(IR15) 동안에 발표자, 토론자, 참여자가 트위터에 올린 메시지를 데이터로 하여 분석한 [10]을 예로 들어 보자. <Table 1>은 [10]에서 기본 문제가 어떻게 변형·응용되었는지를 보여준다. 첫째, ①번의 ‘사람들’은 ‘트위터 사용자’로 구체화되었고, ‘무엇’은 ‘트윗 메시지의 내용’으로 수정되었다. 둘째, ②번의 ‘누가’는 ①번처럼 ‘트위터 사용자’에 해당하며, ‘무엇’은 트윗을 올리면서 인용한 ‘웹사이트 주소’(URL)이다. 여기에서 주체는 사람인 트위터 이용자뿐만 아니라 사물인 트윗 메시지도 가능하다. 따라서 ⑤번의 먼저 나오는 ‘무엇’은 ‘트윗 메시지’이고 뒤에 나오는 ‘무엇’은 그 트윗에서 인용한 ‘웹사이트 주소’(URL)가 된다. 세 번째 문제는 ③번과 ⑥번을 응용하면서 사회네트워크분석(social network analysis: SNA)에서 이용되는 전문 용어를 채택하였다. SNA에서 ‘유력자’는 특정 집단에서 다른 사람들과 가장 많은 관계를 맺은 개인이다. 그리고 트위터에서 관계는 전화 통화와 마찬가지로 정보를 보내는 사람과 받는 사람으로 구분될 수 있다.

<Table 1> Applying primary types of research questions to Twitter big data

Basic types	Questions in Twitter research
① What do people talk?	How often do the participants of the 2014 Internet Research conference post tweets related to conference location? What are those tweets about?
② Who can see what and pays attention to what?	What URLs are linked from tweets posted by conference participants?
⑤ What can link to what?	
③ Who can reply to whom?	Who are influentials in the conference network? What are their profiles?
⑥ Who can link to whom?	

* Basic types are slightly modified by the authors based on Hansen et al. (2011)

두 번째로 페이스북 팬 페이지 데이터를 분석한 [11]을 통해 기본 유형이 어떻게 변형되었는지 그 과정을 검토해 보자. <Table 2>에서 보듯이, 첫 번째 연구문제에서 말하는 ‘댓글’은 ⑤번의 ‘무엇’에 해당한다. 동시 댓글이란 페이스북 팬 페이지에 올라온 특정한 포스트(posts)에 공통으로 남겨진 대답들(comments)을 뜻한다. 동시 댓글 네트워크의 구조는 [Fig. 1]과 같다. 두 번째 문제는 첫 번째 질문과 비교하면, 분석 단위가 ‘무엇’에서 ‘누가’로 바뀌었다. 사물인 ‘댓글’이 아니라 해당 댓글을 남긴 ‘사람’으로 초점이 변화했으므로 ⑥번의 변형이다. 나아가 사회과학 연구의 오랜 관심사인 성별 간 차이 검토를 목적으로 설정하였다. 그렇지만 [11]에서는 [10]과는 다르게 ③번은 응용되지 않았다. 왜냐하면 동시 댓글은 전화 통화와 다르게 송신자와 수신자가 존재하지 않기 때문이다. 즉, [Fig. 1]에서 보듯이 방향성이 없는 구조이다.



[Fig. 1] Visualization of co-commenting network on Facebook

<Table 2> Applying primary types of research questions to Facebook big data

Basic types	Questions in Facebook research
⑤ What can link to what?	What are the characteristics of a co-commenter network represented on fanpages of Daegu-Gyeongbuk's Mayor and Governor candidates?
⑥ Who can link to whom?	How different are the fanpage activities of co-commenters in terms of their gender?

세 번째로, 트위터와 페이스북에 이어서 유튜브 데이터를 이용한 [12]의 강남스타일 사례 분석을 보자. 이 연구는 단행본에 포함된 장(chapter)이어서 학술지 논문처럼 문제를 명시적으로 설정하지는 않았다. 이에 기본 유형에 상응하는 연구문제를 <Table 3>과 같이 정리하였다. ①번의 ‘사람들’은 강남스타일 공식 동영상에 대해 댓글을 남긴 ‘사람들’이다. 그리고 ①번의 ‘무엇들’에 해당하는 것은 댓글에서 가장 많이 언급된 단어들이다. 이 단어들은 [12]에 네트워크 그림으로 시각화되어 있다. ②번의 ‘무엇’은 채널에 방점을 두었기에, 메시지에 초점을 둔 ①번의 ‘무엇’과는 다르다. ②번의 ‘무엇’에서 [12]는 댓글을 남긴 사람들이 유튜브에서 구독하는 ‘채널’들을 조사한 이후에, 그 채널들이 얼마나 많이 중복되었는지를 조사하였다. 즉 강남스타일에 댓글을 남긴 네티즌이 공통적으로 주목하는 유튜브 인기 채널이 ‘무엇’인가를 추적했다. ③번의 ‘누가’는 쉽게 알 수 있다. ‘대답’은 댓글을 주거나 받는 것으로, 댓글을 남긴 사람들의 계정은 동심원으로 표시되고 이들을 연결하는 선의 끝에 화살표가 표시된 것은 방향성을 표시한다. ④번의 ‘콘텐츠’는 댓글에서 많이 사용된 ‘어휘’이다. 그리고 지속성을 측정하기 위해서 특정 ‘어휘’가 8월, 9월, 10월에 수집된 댓글에서 소멸하지 않고 등장하는 빈도를 계산하였다. 나아가 ‘콘텐츠’를 어휘 차원에서 확장하여 개별 ‘댓글’로 간주하여 월별 감성 분석을 실시하였다. ⑤번은 [12]에서 찾기 힘들다. 마지막 ⑥번은 ‘누가’-‘누구와’의 관계에서 방향이 없다는 점에서 ③번과 큰 차이가 있다. 따라서 이 경우에는 선의 끝에 화살표가 없다. [12]에서는 댓글을 남긴 사람들이 유튜브에서 구독하는 채널이 같을수록 강한 관계를 맺는 것으로 변형되었다.

<Table 3> Applying primary types of research questions to Youtube big data

Basic types	Questions in Youtube research
① What do people talk?	What are the demographic profiles of repliers? What word occurs frequently with what?
② Who can see what and pays attention to what?	What channels do repliers subscribe to in common?
③ Who can reply to whom?	What is the communication structure of a reply-to network?
④ How long is content visible and how wide spreadable?	What word in replies has been salient and disappears over time? What sentiment has been expressed in replies over time?
⑥ Who can link to whom?	What is the communication structure of a co-subscription network?

3. 기본 유형에 따른 데이터 형태

앞에서 예로 제시한 연구들을 자세히 살펴보면, 분석을 위한 소셜 빅데이터의 기본 구성은 모두 행렬 즉 매트릭스(matrix)로 되어 있음을 알 수 있다. <Table 4>에서 보듯이 여섯 가지 기본 유형에 대해 매트릭스의 행과 열에 들어가는 항목이 조금 다르기는 하지만, 데이터는 본질적으로 쌍(dyad)으로 구성된다. ①, ②, ④번은 행렬 속 항목의 이름이 서로 다르고, ③, ⑤, ⑥번은 서로 같다. 매트릭스 데이터를 다루는 SNA에서 ①, ②, ④번은 2-모드 혹은 제휴(affiliation) 매트릭스, ③, ⑤, ⑥번은 1-모드 혹은 인접(adjacency 또는 proximity) 매트릭스로 불려진다. 행렬 속 항목에 대해서는 노드(Node) 혹은 꼭지점(Vertex)이라고 불려진다. 1-모드에서 ③, ⑤번과 같이 보내는 쪽과 받는 상대방의 방향을 구분할 필요가 있으면 비대칭적(asymmetric) 매트릭스라 하고, ⑥번과 같이 방향을 나누는 것이 큰 의미가 없으면 대칭적(symmetric) 매트릭스라고 한다.

<Table 4> Matrix format according to primary types of questions

Basic types	Who 1	Who 2	...	Who n
① What do people talk?	Who 1	Who 2	...	Who n

Basic types	Who 1	Who 2	...	Who n
② Who can see what and pays attention to what?	Who 1	Who 2	...	Who n

Basic types	To whom 1	To whom 2	...	To whom n
③ Who can reply to whom?	Who 1	Who 2	...	Who n

Basic types	Time 1	Time 2	...	Time n
④ How long is content visible and how wide spreadable?	Content 1	Content 2	...	Content n

Basic types	What 1	What 2	...	What n
⑤ What can link to what?	What 1	What 2	...	What n

Basic types	With whom 1	With whom 2	...	With whom n
⑥ Who can link to whom?	Who 1	Who 2	...	Who n

4. 통계적 제 문제

소셜 빅데이터 학자들 사이에서 받아들여지는 보편적인 가설은, 현실의 정치·사회·경제·문화적 관계는 인터넷과 SNS를 통한 의사소통 관계에 배태될 수 있다는 것이다[13]. 스마트폰의 대중화는 온라인 관계에서 시작되어 오프라인으로 확장되는 O2O(online to offline) 현상을 가져오기도 했다. 그래서 SNS에서 상호 간 관심을 자주 보이고 댓글을 나누면 오프라인에서 관계를 갖는 경향이 있을 것이라고 가정할 수도 있다. 이 때 미래의 권력이나 비즈니스 관계를 예측하기 위해서는 이러한 SNS 이용자들의 의사소통 관계에 대한 가설을 통계적으로 검증하여야 한다. 그런데 어떻게 하면 이런 유형의 가설을 검증할 수 있을까? <Table 4>에서 볼 수 있듯이, 소셜 빅데이터

의 기본 구성은 매트릭스 형태로 일반적인 통계분석에서의 데이터와는 형태가 다르다. 뿐만 아니라 SNS를 통하여 수집되는 소셜네트워크 데이터는 대부분 모집단으로부터의 무작위 표본(random sample)도 아니며 각 개별 관측값은 상호의존적(interdependent)이다. 따라서 일반적인 추론통계의 방법들을 매트릭스 형태의 데이터에 직접 적용할 수는 없다[14]. 따라서 소셜 빅데이터에 대해 통계적 유의성(significance) 검정 등을 하기 위해서는 별도의 검정방법이 요구된다. 본 연구에서는 [15]와 [16] 등을 통하여 소셜 빅데이터 분석과 관계된 통계적인 제 문제에 대해서 고찰해 보도록 하겠다. 소셜네트워크 데이터처럼 관측값들이 독립성 가정을 충족시키지 못하는 경우 검정통계량(test statistic)을 계산하기 위해서는 표준오차를 추정할 수 있는 부트스트랩(bootstrap) 등의 방법을 활용한 퍼뮤테이션(permutation) 검정을 수행한다. 퍼뮤테이션 검정은 귀무가설이 옳다는 가정 하에서 부트스트랩 방법을 통해 무수히 많은 표본을 추출하고 이들 표본들의 통계량으로 구성된 표본분포를 토대로 통계적 유의성 검정을 수행하는 방법이다. 퍼뮤테이션은 재배열을 의미하는데, 소셜네트워크 관점에서 본다면 재배열 되는 데이터는 네트워크 데이터인 매트릭스이다. 이와 같은 퍼뮤테이션 개념을 바탕으로 하는 통계적 유의성 검정방법은 하나의 네트워크 내에서의 관계뿐만 아니라 두 개 이상의 여러 네트워크 간의 관계를 검정하는 데에도 활용될 수 있다[14]. 네트워크 데이터를 활용한 통계적 검정을 일반적인 통계검정에 대비하여 정리해 보면 <Table 6>과 같은데, 이 중에서 몇 가지만 예를 들어 살펴 보겠다. 먼저 네트워크 데이터에 있어서 평균의 개념은 이진(binary) 데이터인 경우는 연결 관계의 비율 즉 밀도(density)를 의미하고 계량(valued) 데이터인 경우에는 관계의 평균 강도를 의미한다. 따라서 네트워크 데이터에 대한 일 표본 평균검정의 귀무가설은 ‘ H_0 :네트워크의 밀도가 0이다’와 같은 형태로 설정된다. 다음의 <Table 5>는 UciNet6을 활용하여, 국내 25개 국립대학교 도서관 간의 네트워크 자료를 대상으로 수행된 일 표본 평균검정의 결과를 설명하기 위한 예이다. 검정 결과 일반적인 통계검정에서의 표준오차는 0.0233인데 반하여 퍼뮤테이션 검정에서의 표준오차는 0.0601로 약 2.6배 더 크다. 그리고 퍼뮤테이션 검정에서의 검정통계량은

$2.163(= \frac{0.13}{0.0601})$ 이고, 이에 따른 양측검정의 유의확률(p-value)은 0.0328임을 알 수 있다. 따라서 ‘ H_0 :네트워크의 밀도가 0이다’라는 귀무가설은 유의수준 5%에서 기각된다.

<Table 5> Result of one sample average-test using UciNet6

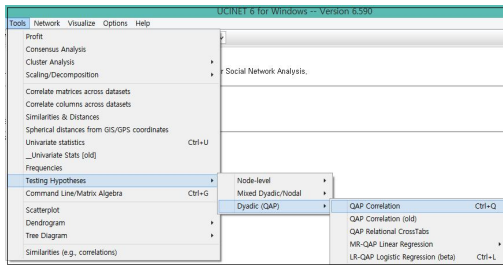
COMPARE DENSITY W/ HYPOTHEZIZED VALUE	

Parameter value is:	0.0000
Density of Library is:	0.1300
Difference is:	0.1300
Variance of ties for Library:	0.0998
Classical estimate of SE:	0.0223
Number of bootstrap samples:	5000
Estimates standard error for density of Library:	0.0601
z-score:	2.163
average bootstrap density:	0.1198
Proportion of absolute difference as large as observed:	0.0328
Proportion of difference as large as observed:	0.0328
Proportion of difference as small as observed:	0.9682

<Table 6> Comparing statistical significance tests

Statistical data in general	Social network data	Significance test for network data
One sample t-test	One sample average-test	Comparing density (or average value in a weighted matrix) of a given network with parameter value hypothesized under investigation
Two-sample t-test	Two-sample average-test	Comparing two networks composed of the same nodes in terms of differences in the pairs of density (or average) values
Independent-samples t-test	Two-group average test	Testing for differences between averages from several metrics (e.g., degree, betweenness, eigenvector, reciprocity, clustering coefficient, etc.) of actors in single network
Analysis of covariance (ANCOVA)	Multiple-group average test	Same with the above but more than three groups are tested
Correlation analysis	QAP correlation analysis	Comparing structural similarity between two networks using permutation test
Regression analysis	QAP regression analysis	Doing regression in QAP by setting one network as independent variable and the other network as dependent one

다음으로 동일한 행위자(actor)들로 구성된 네트워크에서 행위자들 간의 관계가 두 네트워크에 걸쳐 서로 관련이 있는지를 검정하는 즉, 네트워크 자료 간 상관성이 있는지를 검정하는 예를 살펴보자. 이를 위해서는 [Fig. 2]를 활용한 QAP(quadratic assignment procedure) 상관분석을 수행해야 하는데, 두 매트릭스 간의 상관관계의 정도는 피어스의 적률상관계수를 활용한다. 다만 통계적 유의성(statistical significance) 검정을 수행하는 과정에서 퍼뮤테이션 검정의 개념을 활용 한다는 점이 다르며, 따라서 일반적인 상관분석에서와는 다른 유의확률이 도출된다. 이 과정에서 수행되는 퍼뮤테이션 횟수에 따라 유의확률은 달라질 수 있으나, 퍼뮤테이션 횟수가 커질수록 유의확률의 변동성은 작아지게 된다. 나머지 검정을 포함한 보다 자세한 사항은 [17]과 [14] 등에 잘 나타나 있다.



[Fig. 2] Ucinet menu for QAP correlation analysis

5. 결론

최근 페이스북, 싸이월드 그리고 트위터와 같은 소셜 네트워크 사이트(SNS)가 관심을 받으면서 소셜 네트워킹 열풍이 불고 있다고 해도 과언이 아니다. 이와 함께 학계에서는 사람들이 소셜 미디어로부터 생산되는 빅데이터를 활용하여 인간관계를 어떻게 증진하고 정치 체제에 어떤 영향을 미치는지 혹은 직업을 구하고 신뢰를 증진하며 새로운 제품을 거래하는데 소셜 미디어가 어떻게 이용될 수 있는지에 대해 많은 연구를 수행하고 있다[18, 19]. 그런데 연구라는 작업은 증거가 없는 상식을 체계적·구체적·논리적인 방법으로 증거를 확인하여 이론을 정립해 주는 작업이기에 합당한 연구방법이 있어야 한다. 특히 네트워킹 과정에서 만들어지는 자료는 매트

릭스 형태의 자료가 대부분으로 기존의 통계분석에서 사용되는 자료와는 많은 차이가 있다. 따라서 자료의 형태에 부합되는 연구방법과 함께 올바른 연구문제 설정이 이루어져야 하며, 나아가 이를 토대로 한 타당한 통계분석이 적용되어야 한다. 이에 본 연구에서는 소셜 빅데이터 분석에 합당한 연구방법론 개발에 앞서, 연구문제의 설정을 위한 질문의 기본 유형을 다음과 같이 6가지로 체계적으로 정리하고 질문의 기본 유형에 따른 데이터 형태에 대해서도 살펴보았다.

- ① 사람들이 무엇을 얘기하는가?
- ② 누가 무엇을 보고 주목하는가?
- ③ 누가 누구에게 대답하는가?
- ④ 콘텐츠가 얼마나 지속되는가?
- ⑤ 무엇이 무엇에 연결되어 있는가?
- ⑥ 누가 누구와 관계를 맺고 있는가?

또한 SNS로부터 수집되는 데이터는 일반적인 통계검정에서 요구되는 기본 가정들을 만족하지 못하기에 별도의 검정 방법이 필요하다. 이에 본 연구에서는 관측값들이 독립성을 만족하지 못해서 모집단의 이론적 확률분포를 알 수 없을 때 사용되는 부트스트랩 방법을 활용한 퍼뮤테이션 검정에 대해서 고찰해 보았다. 퍼뮤테이션 검정은 비모수적(nonparametric)인 기법으로 소셜빅데이터 관점에서 재배열되는 데이터는 네트워크 데이터 매트릭스이다. 재배열이란 그래프의 구조는 동일하나 행과 열의 노드는 달라지게 됨을 의미한다.

본 연구에서 다룬 소셜 네트워크 빅데이터 분석과 관련된 통계적인 제 문제는 향후 관련 연구자들이 데이터 유형에 맞는 올바른 연구문제를 수립하고 분석함으로써 타당한 정보를 도출하는데 도움이 될 것으로 사료된다.

REFERENCES

- [1] T. M. Song, "Efficient utilizing of big data on health & welfare." Health and Welfare Policy Forum, Vol. 193, pp.68-76, 2012.
- [2] G. H. Lee, Y. K. Ham, Y. D. Kim, J. H. Lee, J. H. Won, "Understanding big data", KNOU Press, 2015.
- [3] S. Rheem, "Smart disaster management strategies utilizing dig data" Korean Review of Crisis &

- Emergency Management, Vol. 10, pp.23-43, 2014.
- [4] J. S. Kim, S. Jin, “A study on the application of opinion mining based on big data”. Journal of the Korean Data Analysis Society, Vol. 15, pp.101-113, 2013.
- [5] J. W. Kim, “Big data’s effect on its practical use for box-office success of films: Focus on the process of management of opinion mining on <Roaring Currents>”. Journal of the Korean Entertainment Industry Association, Vol. 9, pp.81-90, 2015.
- [6] T. J. Seong, K. J. Si, “Research methodology”, Hakjisa, 2007.
- [7] K. W. Kim, “Lecture of methodology”, Pakyoungsa, 1999.
- [8] M. Felt “Social media and the social sciences: How researchers employ Big Data analytics”. Big Data & Society, DOI: 10.1177/2053951716645828, 2016.
- [9] D. L. Hansen, B. Shneiderman, M. A. Smith, “Analyzing Social Media Networks with NodeXL: Insights from a Connected World”, MA: Morgan Kaufmann Publishers, 2011.
- [10] M. K. Lee, J. W. Park, H. W. Park, M. Smith, “Network analysis of the international conference utilizing social media: The case of internet research conference (IR15)”, Tourism Research, Vol. 30, pp.223-246, 2015.
- [11] J. W. Park, H. W. Park, “Social network analysis among Facebook fanpage co-commenters: Daegu-Gyeongbuk’s Mayor-Governor candidates.” Journal of the Korean Data Analysis Society, Vol. 16, pp.3243-3253, 2014.
- [12] J. Y. Park, W. W. Xu, H. W. Park, “Analyzing Youtube big data of Gangnam style”, Kim, S. T. (ed), Communication research in big data era, Yulgokbook Publishing Co., 2015.
- [13] G. A. Barnett, M. S. Lee, K. Jiang, H. W. Park, “The flow of international students from a macro perspective: A network analysis”, Journal of Comparative and International Education, Vol. 46, pp.533-555, 2016.
- [14] K. Y. Kwahk, “Social network analysis”, Chong-Ram, 2014.
- [15] S. P. Borgatti, M. G. Everette, J. C. Johnson, “Analyzing social networks”, Sage Publications Inc., 2013.
- [16] C. Prell, “Social network analysis: A handbook”, Sage Publications Inc., 2012.
- [17] L. J. Hubert, “Assignment methods in combinatorial data analysis”, Marcel Dekker, 1997.
- [18] L. S. Kim, “Convergence of information technology and corporate strategy”, Journal of the Korea Convergence Society, Vol. 6, pp.17-26, 2015.
- [19] G. S. Ryu, “Development of educational model for ICT-based convergence expert”, Journal of the Korea Convergence Society, Vol. 6, pp.75-80, 2015.

박 한 우(Park, Han Woo)



- 1995년 2월 : 한국외국어대학교 신문방송학과(정치학사)
- 1997년 2월 : 서울대학교 신문학과(언론정보학 석사)
- 2002년 6월 : 뉴욕주립대학교 커뮤니케이션학과(언론정보학 박사)
- 2003년 3월 ~ 현재 : 영남대학교(언론정보학과 교수)

- 관심분야 : 빅데이터, 웹보메트릭스
- E-Mail : hanpark@ynu.ac.kr

최 경 호(Choi, Kyoung ho)



- 1985년 2월 : 전북대학교 전산통계학과(이학사)
- 1995년 2월 : 서울대학교 계산통계학과(이학박사)
- 1993년 3월 ~ 현재 : 전주대학교(기초의과학과 교수)
- 관심분야 : 통계조사, 사회연결망분석, 데이터 테크놀로지

- E-Mail : ckh414@jj.ac.kr