

Research topic analysis of the European Sport Management Quarterly: Topic modeling with Latent Dirichlet Allocation(LDA)

Mihwa Choi¹, Hyungil Kwon^{1*}, Juhae Baek¹, & Do Young Pyun²

¹Chung-Ang University & ²Loughborough University

[Purpose] The purpose of this study was to analyze the research topics of the articles which were published through European Sport Management Quarterly(ESMQ) from 2009 to 2018. The prior topic analysis studies of the ESMQ classified topics based on the key words using NASPE-NASSM SMPs categories. Therefore they couldn't fully reflect the content of the articles. **[Methods]** The topic modeling of the current study was conducted with the Latent Dirichlet Allocation(LDA) which generates topics based on the word usage in the article. A total of 265 articles were converted from 'pdf' format to 'txt' ANSI format for topic modeling analysis. The whole topic modeling process was done using R program and the model was set to generate 10 topics from the article. **[Results]** The 3 sport management experts were hired to label the name of the topics and the name of the topics are as follow : (1) Impact of mega sport event, (2) Cause-related marketing, (3) Factors affecting the results of the competition, (4) Managing sport organization, (5) European sport leagues, (6) Strategic management, (7) Sport economics, (8) Sport in communities, (9) Sport consumers, (10) Elite sports. It is not quite possible to compare the results of the current topic modeling results with the previous ones because of the methodological differences. However, even though the standards are different, Sport marketing topic showed the largest growth among the 10 topics extracted. **[Conclusions]** This study used the LDA probabilistic algorithm to analyze research topics, which made the analyses more objective and wholistic. However, the insights of the researchers were still needed to interpret and labeling the topics.

Key words: text mining, topic modeling, LDA, content analysis, European Sport Management Quarterly

서론

스포츠 산업의 발달과 함께 스포츠 경영의 학문적 발전을 촉진하기 위하여 다양한 학회가 형성되었고 그 학회에서는 연구의 발전을 위한 노력의 일환으로 학회지를 발간하기 시작하였다. 세계 스포츠 산업의 한 축으로 인식되고 있는 유럽에서는 1992년에 처음으로 학회(European

Association for Sport Management, EASM)가 설립되었다. 이후 1994년부터 EASM에 의해 European Journal for Sport Management(EJSM)가 발간되기 시작하였고 그 저널은 현재 European Sport Management Quarterly(ESMQ)로 명칭이 바뀌었다. ESMQ는 유럽과 관련된 연구뿐만 아니라 스포츠 경영에 대한 여러 측면을 서로 다른 문화적, 지역적 관점으로 연구한다는 학문적 철학을 보유하고 있다(ESMQ, 2019). 학회지가 보유한 철학을 바탕으로 이론 개발 및 경험이 적용된 논문을 발행함으로써 스포츠 경영 분야에 새로운 지식을 제공하고 있다.

논문 투고일 : 2019. 09. 25.

논문 수정일 : 2019. 10. 27.

게재 확정일 : 2019. 11. 05.

* 교신저자 : 권형일(hkwon@cau.ac.kr).

ESMQ는 스포츠 경영 전문 학술지로서 인정받고 있으며 학술지에 게재되는 논문들이 스포츠 경영학 분야를 잘 반영하고 있는 것으로 평가되고 있다. 이는 1,565개의 경영관련 학술지를 분석한 Academic Journal Guide 2018에서 확인 할 수 있다(CABS, 2018). 발표된 자료에 의하면 ESMQ에서 발행되어진 논문들이 스포츠 경영 전문 학술지라는 취지와 부합하다고 평가 되었고, 인용 지수와 심사자의 평가에 따라 가장 높은 등급의 스포츠 경영 학술지로 구분되었다. 이러한 학술지를 통해 발표된 논문들은 학문의 발전에 기여를 하는 것으로 인식되어지기 때문에(Van Doren & Heit, 1973) ESMQ에 어떠한 연구 논문들이 발표되고 있는가에 대한 분석이 여러 차례 시도되었다(Ciomaga, 2013; Pitts, Danylchuk & Quarterman, 2014; Shilbury, 2011).

연구자들은 ESMQ에서 발간된 논문들에서 자주 참고되고 있는 학술지가 어느 것인지(Shilbury, 2011), 연구자들의 관심의 초점이 되고 있는 연구 영역은 무엇인지(Ciomaga, 2013; Pitts et al., 2014)등을 연구하면서 10년 이상의 기간에 걸쳐 생성된 자료를 사용하였다. 학술지를 통해 10년 이상의 기간 동안 발행한 논문들을 분석해야 하는 연구들은 방대한 양의 문헌을 분석해야 한다는 어려움을 갖고 있다. 그렇다보니 숙련된 전문 코디라 할지라도 논문의 전문을 모두 읽고 주제를 분류하고 이 결과를 입력하기에는 시간적, 재정적 어려움이 따를 수밖에 없다. 이러한 이유로 사전 연구에서는 논문의 전문을 사용하기 보다는 주제어(key word)나, 참고문헌, 제목 등을 사용하여 분석하였다. Shilbury(2011)의 경우 ESMQ에서 발간된 논문들에서 자주 참고되고 있는 학술지가 어느 것인지를 연구하였기 때문에 참고문헌을 분석 자료로 사용하는 것이 당연한 선택이라고 볼 수 있다. 하지만 발행된 논문들에서 주된 연구 주제를 찾는데 논문의 전문을 사용하지 않고 참고문헌만을 바탕으로 하거나(Ciomaga, 2013), 주제어(key word)를 바탕으로 하여 분석(Pitts et al., 2014) 하는 것은 선행 연구들이 갖고 있는 약점이라고 할 수 있다.

ESMQ의 주제분석을 다룬 선행연구의 또 다른 약점은 자료 분석 시에 적용한 분류 기준에 대한 부분이다. 예를 들어 Pitts et al.(2014)의 연구에서는 NASPE-NASSM Sport Management Program Standards(SMPS, Sport

Management Program Review Council, 2000)의 10개 주제를 분류의 기준으로 삼았는데 NASPE-NASSM SMPS는 스포츠경영학이라는 대학 전공 안에서 어떠한 교과목들이 포함되어야 하는지에 대한 기준일 뿐이며, 이 기준이 스포츠경영학에서 이루어지고 있는 연구의 주제 분류 기준은 아닌 것이다. 또한 Pitts et al.(2014)의 연구에서는 Sport management education이 연구의 주제 분류 항목으로 포함이 되어있다. 이 연구들은 스포츠경영학을 어떻게 가르쳐야 하는지에 대한 연구로 스포츠경영학의 내용학이 아닌 교육의 방법적인 부분을 포함한다.

선행연구에서 이러한 약점들이 나타나는 것은 출판된 연구들의 내용을 전체적으로 파악하고 이 연구들을 분류한 것이 아니라, 주제 리스트를 미리 정해놓고 각 연구들의 제목, 초록 혹은 주제어를 바탕으로 각 주제에 할당하는 방식을 사용하기 때문이다. 이러한 선행연구들에서 나타난 약점을 보완하기 위해서는 방대한 양의 논문의 전문 안에 내제된 정보들의 연관성을 파악해 내어 논문들이 가지고 있는 경향에 따라 분류하는 분석방법이 요구된다. 이러한 요구를 충족시키는 분석방법으로 토픽 모델링(topic modeling)이라는 새로운 분석기법이 사용되고 있다. 토픽 모델링은 문서 안에 들어 있는 텍스트들의 의미 구조를 발견하기 위해 사용되는 텍스트 마이닝(text mining) 기법 중에 하나이다. 이 기법은 컴퓨터 알고리즘을 통해 문헌에 등장하는 단어들의 분포를 파악하여 문헌 내에 어떠한 주제가 들어 있으며, 주제들의 비중이 어떠한지 등을 객관적으로 제시해준다(Blei, 2012). 본 연구에서는 ESMQ를 통해 최근 10년 동안 출판된 논문들이 어떠한 주제들을 다루고 있는지를 토픽 모델링을 사용하여 파악하고자 한다.

ESMQ와 유사 저널에 대한 주제 분석 결과 고찰

다수의 스포츠 경영 전문 학술지 중에서 유럽을 대표하는 학술지로 ESMQ가 있다면 북미를 대표하는 학술지로는 Journal of Sport Management(JSM)가 존재한다. ESMQ와 JSM은 스포츠경영학에서 대표적인 저널이면서 각 대륙을 대표하는 학술지로 인정받고 있다. 예를 들어 JSM은 스포츠 경영 학술지들 가운데 가장 친숙한 학술지로 인식되고 있으며 학술지의 품질 평가에서 가장 높

은 점수를 기록했고(Shilbury & Rentschler, 2007) ESMQ는 학회가 설립된 유럽과 타 지역에 존재하는 스포츠 경영과 관련된 학회지 중에서 가장 높은 Impact Factor를 보유하고 있다(Clarivate Analytics, 2019).

이렇다 보니 각 대륙을 대표하는 JSM과 ESMQ를 통해서 어떠한 논문들이 발표되고 있는지에 대한 학자들의 관심이 있어왔다. 그중에서 JSM에 대한 종합적인 주제분석은 Pitts & Pedersen(2005)에 의해 처음으로 시행되었다. 이들은 주제 분류의 대상 논문들을 NASPE-NASSM SMPS에서 제시하는 10개의 주제에 'Sport management education'과 'Others'를 포함하여 12개의 주제로 분류하였다. 이 연구에서는 Management & organizational skills in sport(n=89, 38%)가 가장 많이 연구된 주제인 것으로 나타났다. 이후에 Kwon, Pyun & Lee(2008)에 의해 진행된 JSM에 대한 분석은 8개의 주제로 분류되었는데 이들은 Parkhouse(2005), Parks, Zanger & Quarterman(1998), Slack & Parent(2006)의 문헌을 바탕으로 하위학문분야를 8개로 구성하였다. 이 연구결과 역시 'Sport management'(36.1%)가 연구자들에게 가장 관심을 많이 받았던 주제인 것으로 나타나 Pitts & Pedersen(2005)의 연구 결과와 일관되게 나타났다. 위에서 제시되었듯이, 선행연구에서 주제 분석의 분류 기준이 12개 혹은 8개로 다소 차이가 있기는 하지만 연구자들에게 가장 많은 관심을 받고 있는 주제는 스포츠 경영학 중에서도 스포츠조직에 대한 경영 부분의 주제인 것으로 판단할 수 있다.

Kim & Cho(2013)는 상술된 연구에서 채택한 범주와 다른 범주로 JSM의 연구주제를 분석하였다. 그들의 연구에서는 연구영역, 연구내용, 연구방법 세 가지의 큰 범주를 채택하고 각각의 범주 아래에 하위 범주를 설정하였다. 이 연구는 내용분석(Content analysis) 기법을 사용하였고 국내 학술지 중 한국스포츠산업경영학회지(Korean Journal of Sport Management(KJSM))를 선정하여 비교분석을 실시하였다. 먼저 연구영역에 대한 분석결과를 살펴보면 JSM은 1건(0.4%)의 논문만 할당된 스포츠용품유통업에 KJSM은 31건의 논문이 할당되어 전체 연구의 8.6%를 차지하는 연구영역인 것으로 나타났다. 구체적으로 하위 범주로 구분한 기타스포츠산업(JSM n=119, 47.8% / KJSM n=83, 23.1%), 스포츠경기업

(JSM n=80, 33.2% / KJSM n=114, 31.7%), 스포츠 시설운영업(JSM n=18, 7.5% / KJSM n=94, 26.1%), 스포츠마케팅(JSM n=7, 2.9% / KJSM n=19, 5.3%), 스포츠정보업(JSM n=14, 5.8% / KJSM n=18, 5.0%) 등에서 JSM과 KJSM 연구자들의 관심이 통계적으로 유의한 차이가 있는 것으로 보고 하였다. 연구내용에 대한 비교분석 결과도 두 학술지간에 통계적으로 유의한 차이가 보고되었다. 예컨대 '윤리적이슈'와 관련된 연구내용이 JSM에서는 36건(14.9%)의 논문이 할당되었지만 KJSM은 단 1건(0.3%)의 논문이 할당되었다. 하지만 두 학술지에서 공통적으로 가장 많은 논문이 할당된 분야 역시 '마케팅 관리'(JSM n=71, 29.5% / KJSM n=239, 66.4%)인 것으로 나타났다.

본 연구의 분석 대상인 ESMQ에 대해 가장 최근에 이루어진 종합적인 주제 분석은 Pitts et al.(2014)에 의해 진행되었다. 이 연구에서는 ESMQ의 전신인 EJSM에서 2008년부터 발행한 85개의 논문과 ESMQ에서 2012년까지 발행한 222개의 논문을 포함하여 총 307개의 논문이 분석대상으로 선정되었다. 주제분석을 위해 총괄적 내용분석 방법(Summative content analysis)을 사용하였다. 총괄적 내용분석 방법은 논문에서 추출한 주제어를 사용하는데, ESMQ에는 2011년 2호 이후부터 주제어가 명시되어 있다. 따라서 주제어가 제시되어있지 않은 논문마다 4명의 코더를 통한 주제어 선정 과정이 선행되었다. 하지만 코더들에 의해 어떠한 과정을 통해 주제어가 선정되었는지 설명이 되어 있지 않은데다 선정된 주제어들도 제시 되지 않아 이에 대한 타당성의 약점이 존재한다. Pitts et al.(2014)의 연구에서도 NASPE-NASSM SMPS에 제시된 10개의 주제가 주제분석의 분류 기본틀이 되었고 여기에 Sport management education 항목을 추가하여 총 11가지의 분류기준이 설정되었다. 분석결과는 EJSM과 ESMQ를 나누어서 제시되었는데 공통적으로 Management & organizational skills in sport (EJSM n=32, 37.6% / ESMQ n=53, 23.8%)로 분류되는 논문들이 가장 많은 것으로 나타났다. 이러한 결과는 JSM을 대상으로 분석하여 Management & organizational skills in sport(n=89, 38%)가 가장 많이 연구된 주제인 것으로 보고한 Pitts & Pedersen(2005)의 연구 결과와 Sport management(36.1%)가 가장 많은 연구 주제인

것으로 보고한 Kwon et al.(2008)의 연구 결과와 같음을 알 수 있다. 다음으로 Sport Business in the Social Context(EJSM n=19, 22.3% / ESMQ n=66, 29.3%), Sport Marketing(EJSM n=15, 17.6% / ESMQ n=34, 15.3%), Sport Economics(EJSM n=8, 9.4% / ESMQ n=21, 9.5%) 등 총 네 개의 주제가 다른 일곱 개의 주제에 비해 높은 비율의 연구주제인 것으로 분류되었다. 즉 시기적으로 볼 때 학술지가 창간된 후 20여 년간 연구자들의 주된 연구 관심사는 몇몇의 특정 주제에 머무르고 있었다고 볼 수 있다.

Pitts et al.(2014)이 추가로 선정한 주제인 Sport Management Education은 EJSM(n=6, 7.0%)에서는 다섯 번째로 높은 비율인 것으로 나타났으나 ESMQ(n=5, 2.3%)에서는 절대 수는 많이 감소하지 않았지만 전체 대비 비율로는 많이 감소한 것을 알 수 있다. 이후의 순위인 Communication in Sport(EJSM n=2, 0.2% / ESMQ n=4, 1.8%), Ethics in Sport Management(EJSM n=1, 0.1% / ESMQ n=5, 2.3%), Sport Law(EJSM n=1, 0.1% / ESMQ n=8, 3.6%), Governance in Sport(EJSM n=1, 0.1% / ESMQ n=20, 9.0%), Finance in Sport(EJSM n=0, 0.0% / ESMQ n=4, 1.8%), Field Experience(EJSM n=0, 0.0% / ESMQ n=2, 0.9%) 역시 상위 순위로 나타난 네 개의 주제에 비해 현저히 적은 수의 논문이 할당된 것으로 보고되었다. 그나마 Governance in Sport는 EJSM에는 n=1, 0.1%이었으나 ESMQ에서는 n=20, 9.0%로 향상되어 ESMQ만으로 본다면 다섯 번째로 많은 연구가 이루어지고 있는 주제임을 알 수 있다.

연구방법

자료수집

본 연구에서는 2009년부터 2018년까지 ESMQ에서 발행된 논문들이 사용되었다. 이 기간 동안 발행된 각 호에는 Editorial, Keynote address 등 다양한 유형의 글들도 존재한다. 하지만 본 연구의 목적이 논문들의 연구 주제를 파악하는데 있기 때문에 Article로 구분된 연구논

문들만 사용되었다. 온라인 데이터베이스를 통해 전체 265개의 논문이 PDF 형식의 파일로 수집되었다. 자료를 PDF 형식의 파일로 저장할 때 논문이 발행된 권과 호 그리고 저자명을 파일의 이름으로 지정하여 자료가 누락되거나 이중으로 수집되는 것을 방지하였다. 이렇게 수집된 자료는 분석을 위해 R 프로그램을 통해 ANSI 형식의 Text File로 변환되어 사용되었다.

참고문헌만을 이용하여 주제분석을 한 선행연구가(Ciomaga, 2013; Shilbury, 2011) 있다는 것은 참고문헌의 내용이 주제분석에 영향을 줄 수 있다는 것을 의미한다고 판단하여 본 연구에서는 모든 Text File에서 참고문헌을 수작업으로 삭제하였다. 뿐만 아니라 변환된 Text File에서 주제와 관련이 깊은 것으로 보이는 학술지명(European Sport Management Quarterly)과 페이지마다 반복적으로 삽입되어 있는 논문의 제목도 일일이 삭제하여 주제 분석에 영향 미칠 것으로 판단되는 가외적인 요인들을 최대한 차단하였다.

분석절차

본 주제 분석에는 여러 주제가 혼합된 문서 분석에 가장 적절한 방법으로 알려진 Latent Dirichlet Allocation(LDA)기반의 Topic Modeling 방법이 사용되었다. 일반적으로 LDA를 이용한 주제 분석을 위해 연구자의 판단에 따라 stopwords가 적용된다. 하지만 본 연구에서는 학술지에서 보이는 특징을 확인하기 위하여 연구자가 별도로 선정한 stopwords를 적용하지 않았다. 다만 특수문자, 구두점, 숫자, 텍스트간의 여백을 삭제하는 과정을 거쳐 알파벳만을 남겨 두었고 모든 글자를 소문자로 변환하였다. 또한, 명확한 주제 분석의 도출을 위하여 어간이 같고 어미가 틀린 단어들이 다른 단어로 분류되지 않도록 tm_map(docs, stemDocument)을 사용하였다. 이로 인해 주제 분석 결과에 단어의 전체 스펠링이 제시되지 않는다.

토픽 모델링에서 토픽수의 선택은 일종의 렌즈 선택과 유사한 것으로 비유되는데 소수의 토픽수를 지정하여 주제를 망원렌즈처럼 볼 수 있으며 토픽수를 많이 지정하여 세밀히 들여다보는 현미경처럼 분석할 수 있다(DiMaggio, Nag, & Blei, 2013). 이러한 관점으로 볼 때 토픽수의 최적화 시도는 무의미하며 산출되는 토픽의

해석가능성을 고려하여 연구자의 판단에 따라 토픽수를 결정한다(Chang, Gerrish, Wang, Boyd-Graber, & Blei, 2009). 본 연구에서는 선행연구의 주제 분석들 중에서 가장 기본이 되는 주제의 수인 10개를 토픽수로 결정하여 사용하였다.

R 스크립트를 이용한 분석절차에 따라 4개의 CSV파일을 생성하였다. word_freq 파일은 분석한 문서 전체에서 사용빈도가 가장 높은 단어부터 제시해주는 것으로 단어의 사용빈도가 높다는 것은 해당 단어가 주제 분석에 미치는 영향이 크다는 것을 의미한다. TopicToTerms 파일은 추출된 10개의 주제를 대표하는 단어를 주제별로 10개씩 제시해주며 DocsToTopics 파일은 각각의 논문들이 어떤 주제에 해당하는지 구분하여 준다. 그리고 TopicProbabilities 파일은 각각의 논문들이 10개의 주제와 얼마만큼의 관계가 있는지를 보여주는 Gamma 수치를 제시해 준다.

각 주제별 명칭을 정하는 과정에는 JSM의 논문에 대한 주제 분석 경험이 있는 스포츠경영 국내외 교수 2명과 박사과정의 2인이 참여하였다. 명명의 과정에서는 각 주제별로 제시된 대표 단어 10개와 Gamma수치가 가장 높은 상위 5개의 논문의 주제가 사용되었다. 먼저 박사과정의 2인이 Gamma와 빈도수가 높은 단어를 기초로 하여 주제에 대한 명명을 하였고 이를 국내 스포츠 경영 교수가 수정을 하였다. 이 결과는 국외의 스포츠경영 교수에게 전달되어 최종으로 주제에 대한 명명이 이루어졌다.

연구결과

자료의 일반정보

ESMQ는 2009년부터 2015년까지 평균 24.3편의 논문을 발행하다가 2016년부터는 평균 32편이 발행되었다. 발행된 개별 논문들이 어떠한 주제로 나뉘지는지 분류하였는데 연구가 특정한 주제로 집중되는 현상은 발견되지 않았다. 다만 2010년에는 9편의 논문이 Topic 4로 분류되어 그 해 출판되어진 25편의 논문에서 연구자들에게 가장 관심(36.0%) 받은 연구 주제인 것으로 나타났다. Topic 4는 연구자에게 꾸준히 관심을 받은 주제이며 Topic 3과 Topic 8, Topic 9 역시 10년간 단 한 번도 빠지지 않고 출판되어온 주제이다(Table 1).

자료의 주제분석을 위한 과정 중에 제일 먼저 생성되는 word_freq 파일에는 전체 분석대상 논문에서 사용된 단어들의 빈도수를 제시해준다(Table 2). 예를 들어 제일 높은 빈도수를 나타내고 있는 sport는 문서 전체에서 sport가 25,993회 포착된 것이 아니라 sport, sports, sporty 등과 같이 어미가 동일한 단어들을 한 단어로 인식하여 사용 빈도수가 제시되었다. 이것은 제시된 모든 단어가 가지는 공통점이다.

다른 파일은 10개의 주제로 논문들을 분류할 때 어떠한 기준으로 분류하였는지에 대한 근거를 단어로 제시하여준다(Table 3). LDA 알고리즘은 이 단어들이 가지는 개별적인 의미가 분류 기준이 되는 것이 아니라 단어들이

Table 1. Number of the published articles of ESMQ per year and per topic

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Topic 1	2	2	4	5	3	2	7	1	1	-
Topic 2	-	1	-	2	3	3	2	7	3	2
Topic 3	5	1	1	2	1	1	2	2	2	2
Topic 4	5	9	3	3	4	4	5	2	5	3
Topic 5	1	3	1	2	2	3	-	6	5	4
Topic 6	3	3	2	-	2	3	3	2	1	4
Topic 7	4	2	3	3	-	2	1	3	2	1
Topic 8	1	3	3	2	6	1	1	4	2	3
Topic 9	1	1	3	3	4	6	1	3	5	9
Topic 10	1	-	1	3	-	1	3	2	5	4
	23	25	21	25	25	26	25	32	31	32

Table 2. Word frequency

word	frequency	word	frequency	word	frequency	word	frequency	word	frequency
sport	25,993	model	4,460	data	3,166	find	2,561	inform	2,304
team	8,535	level	4,445	organis	3,055	involv	2,559	interest	2,251
event	7,821	valu	4,204	process	3,032	high	2,557	competit	2,220
studi	7,698	athlet	3,937	increas	3,001	influenc	2,556	exampl	2,210
particip	6,712	posit	3,912	footbal	2,893	ommuniti	2,525	mean	2,199
research	6,099	group	3,869	olymp	2,832	specif	2,523	profession	2,196
manag	5,902	import	3,671	includ	2,831	member	2,514	chang	2,194
club	5,654	may	3,653	market	2,810	structur	2,485	identifi	2,191
use	5,639	activ	3,650	relationship	2,807	nation	2,481	will	2,185
can	5,255	perform	3,649	govern	2,793	associ	2,471	suggest	2,169
organ	5,049	fan	3,539	base	2,789	work	2,471	context	2,143
social	4,981	volunt	3,534	consum	2,760	measur	2,470	perceiv	2,127
differ	4,946	support	3,475	year	2,745	leagu	2,463	follow	2,119
also	4,791	provid	3,462	howev	2,716	role	2,427	toward	2,111
develop	4,767	signific	3,439	player	2,664	brand	2,399	like	2,109
effect	4,696	individu	3,327	tabl	2,653	experi	2,377	term	2,101
game	4,658	factor	3,318	impact	2,649	new	2,367	peopl	2,095
relat	4,646	two	3,293	first	2,605	indic	2,341	need	2,067
one	4,536	variabl	3,281	analysi	2,596	focus	2,336	servic	2,053
result	4,494	time	3,270	within	2,565	number	2,332	attitud	2,050

어떤 단어들과 함께 사용되고 있는가를 바탕으로 주제를 추출한다.

토픽별 명명

TopicProbabilities 파일에 제시된 Gamma 수치가 높을수록 주제와 관계가 높은 논문이다. 따라서 주제별로 Gamma 수치가 높은 상위 5개의 논문을 각각의 주제를 대표하는 논문으로 선정하였다(Table 4-1, 4-2).

Topic 1에는 265개 문서 중 27개 문서(10.2%)가 할당되었으며, 주요 단어로는 'event', 'olympic', 'impact', 'legaci' 등이 선정되었다. 할당된 논문의 내용과 주요 단어를 바탕으로 Topic 1은 "Impact of mega sport event"로 명명하였다.

Topic 2에는 'brand', 'sponsor', 'attitud' 등의 단어가 선정되었다. 선정 단어와 관련 논문 내용의 큰 줄기는 스포츠마케팅으로 그 중에서도 스폰서십 등의 활동이 브랜드에 미치는 영향이 집중적으로 나타났다. 특히 기업이나

브랜드를 사회적 명분이나 이슈에 전략적으로 연계시키는 Cause-related marketing 관련 논문이 두 개나 포함되었다. 이에 따라 스포츠마케팅의 중요 부분인 "Cause-related marketing"으로 Topic 2를 규정하였다.

Topic 3으로 도출된 논문들 중 가장 높은 Gamma 수치를 보인 5개의 논문은 모두 경기의 수행수준에 영향을 줄 수 있는 변인들에 대한 연구로 대표될 수 있었다. 그 요인이 팀내의 요인(i.e., hazing), 제도의 요인(i.e., round robin), 작전(i.e., strategic resting), 심판, match fixing으로 다양하게 제시되었지만 결국 이 요인들은 경기 결과나 경기 수행 수준을 결정짓는 요인들로 분류되어질 수 있었다. 따라서 본 주제는 "Factors affecting the results of the competition"으로 명명 되었다.

Topic 4에는 10개의 토픽 중 가장 많은 문서인 43개(16%)가 할당되었다. 이와 관련된 단어로는 'manag', 'organis' 등 스포츠 경영 연구에서 공통적으로 들어갈 수 있는 단어가 상위에 나타났다. Gamma 수치가 높은 상위 다섯 개의 논문에서는 'culture', 'balancing', 'leadership'

Table 3. Number of assigned articles for a topic and representative words

Topic	n	words
1	27	event, olymp, game, host, impact, resid, social, legaci, citi, local
2	23	brand, sponsor, sponsorship, attitud, consum, toward, team, imag, studi, athlet
3	19	player, team, coach, perform, match, dope, use, stage, haze, fix
4	43	sport, manag, organis, research, process, develop, structur, knowledg, within, can
5	27	team, leagu, footbal, club, season, match, variabl, market, attend, effect
6	23	organ, club, sport, manag, govern, board, member, profit, financi, organiz
7	21	sport, particip, activ, variabl, model, time, age, level, studi, differ
8	26	volunt, social, communiti, sport, group, particip, work, member, team, individu
9	36	sport, fan, team, valu, studi, item, model, research, factor, game
10	20	sport, athlet, elit, perform, develop, nation, coach, countri, polici, support

Table 4-1. Top 5 articles for each topic and the name of the topic

Topic	Gamma	Assigned article
Impact of mega sport event (n=27)	0.84	Legacy perceptions among host and non-host Olympic Games residents: a longitudinal study of the 2010 Vancouver Olympic Games
	0.82	The importance of legacy outcomes for Olympic Games four summer host cities residents' quality of life: 1996-2008
	0.78	The Social Impact of the Tour de France: Comparisons of Residents' Pre- and Post-event Perceptions
	0.72	Legacy of the Beijing Olympic Games: a non-host city perspective
	0.67	Residents' Perceptions of Environmental Impacts of the 2008 Beijing Green Olympic Games
Cause-related marketing (n=23)	0.86	Ending a sponsorship relationship: consumers' responses toward a forced versus a chosen exit
	0.85	The impact of processing athlete transgressions on brand image and purchase intent
	0.78	Exploring the impact of country-of-origin fit and team identification in sports brand evaluation
	0.73	Cause-related marketing in sports: the power of altruism
	0.71	Cause-related sport marketing: an organizing framework and knowledge development opportunities
Factors affecting the results of the competition (n=19)	0.78	A study of efficiency monitoring systems for match-fixing players in the Chinese Professional Baseball League
	0.77	Tie-breaking in round-robin soccer tournaments and its influence on the autonomy of relative rankings: UEFA vs. FIFA regulations
	0.75	Case Law Analysis Regarding High School and Collegiate Liability for Hazing
	0.71	The effect of strategic resting in professional cycling: evidence from the Tour de France and the Vuelta a Espana
	0.63	Payback Calls: A Starting Point for Measuring Basketball Referee Bias and Impact on Team Performance
Managing sport organization (n=43)	0.81	Culture and the Study of Sport Management
	0.79	Embracing Cultural Contexts and Critical Reflexivity: (Re)presenting the Global Sports Industry in Research and Practice
	0.75	Navigating austerity: balancing 'desirability with viability' in a third sector disability sports organisation
	0.75	Creating and leveraging knowledge to promote sport participation: the role of public governing bodies of sport
	0.74	Stakeholder Constructions of Leadership in Intercollegiate Athletics

Table 4-2. Top 5 articles for each topic and the name of the topic

Topic	Gamma	Assigned article
European sport leagues (n=27)	0.88	Fan preferences: one country, two markets and different behaviours
	0.88	Analysing match attendance in the European Rugby Cup: Does uncertainty of outcome matter in a multinational tournament?
	0.85	The lack of balance in the Spanish First Division football league
	0.85	Can sporting success in Norwegian football be predicted from budgeted revenues?
	0.83	The Beckham effect: examining the longitudinal impact of a star performer on league marketing, novelty, and scarcity
Strategic management (n=23)	0.88	Growth determinants of small- and medium-sized fitness enterprises: empirical evidence from Portugal
	0.74	Corporate governance and earnings management in the European football industry
	0.70	Organizational performance of nonprofit and forprofit sport organizations
	0.69	A structural perspective of cross-sector partnerships involving youth sport nonprofit organizations
	0.66	A model of transparency: determinants and implications of transparency for national sport organizations
Sport economics (n=21)	0.92	The Demand for Sports and Recreational Services: Empirical Evidence from Germany
	0.92	Travel Time Spending Behaviour in Recreational Sports: An Econometric Approach with Management Implications
	0.92	The determinants and income elasticities of direct and indirect sports expenditure categories
	0.89	An examination of the determinants of Irish household sports expenditures and the effects of the economic recession
	0.88	Expenditures on Sport Apparel: Creating Consumer Profiles through Interval Regression Modelling
Sport in communities (n=26)	0.78	Sunday afternoon social capital: an ethnographic study of the Southern City Jets Club
	0.76	Creating Communities that Lead to Retention: The Social Worlds and Communities of Umpires
	0.70	Changing roles: applying continuity theory to understanding the transition from playing to volunteering in community sport
	0.68	The Value of the Dark Side: An Insight into the Risks and Benefits of Engaging in Health-compromising Practices from the Perspective of Competitive Bodybuilders
	0.64	'You belonged to something': exploring how fundraising teams add to the social leverage of events
Sport consumers (n=36)	0.92	The relative effects of game outcome and process on fans' media consumption experiences
	0.88	Psychometric Evaluation of the Team Identification Scale among Greek Sport Fans: A Cross-validation Approach
	0.85	Managing dive centres: SCUBA divers' behavioural intentions
	0.85	The effects of emotions on football spectators' satisfaction and behavioural intentions
	0.85	Impact of core and peripheral service quality on consumption behavior of professional team sport spectators as mediated by perceived value
Elite sports (n=20)	0.84	Positioning in Olympic Winter sports: analysing national prioritisation of funding and success in eight nations
	0.84	Global Trade in Sports Goods: International Specialisation of Major Trading Countries
	0.74	A psycho-educational curriculum for sport career transition practitioners: development and evaluation
	0.71	Exploring China's success at the Olympic Games: a competitive advantage approach
	0.69	Coach migration in German high performance sport

등 스포츠조직의 운영과 관련된 내용들이 포함되었다. 이에 따라 본 주제는 "Managing sport organization"으로 명명되었다.

Topic 5에는 'team', 'leagu', 'club' 등의 단어가 상위 에 도출되었으며, 관련 논문에서도 'European Rugby Cup', 'Spanish First Division football', 'Norwegian

football' 등 유럽 국가의 스포츠 리그에 대한 내용이 주를 이루었다. 이와 같은 특성을 반영할 수 있도록 본 주제는 "European sport leagues"로 명명되었다.

Topic 6에는 'organ', 'club', 'govern', 'profit' 등의 단어가 선정되었으며 관련 논문에서도 조직경영에 대한 내용이 주를 이루고 있었다. 또한 다섯 개의 논문이

'earnings', 'performance', 'transparency', 'structure', 'growth' 등의 주제어를 포함하면서 조직 운영의 전략적인 부분을 다루고 있는 측면을 반영하여 "Strategic management"로 명명하였다.

Topic 7에서는 'particip', 'activ' 등의 단어들이 출현 빈도가 높은 단어로 나타났다. 그러나 Gamma 수치가 높은 상위 5개의 논문의 제목에서는 상당히 일관되게 스포츠와 관련된 지출에 대한 내용이 포함되어 있었다. 다섯 개의 논문에서는 레크리에이션에 대한 지출, 스포츠 의류에 대한 지출 등 다양한 지출의 대상이 존재한다. 하지만 같은 주제로 묶인 다른 16개의 논문들을 확인한 결과 개인의 지출에 관련된 내용과 공적 지출에 관한 내용들이 혼재 되어 있는 것으로 나타났다. 따라서 이를 모두 아우를 수 있도록 "Sport economics"으로 주제에 대한 명명을 결정하였다.

Topic 8에는 'volunt', 'social', 'communiti' 등의 단어가 출현빈도가 높은 단어들로 선정되었다. 관련 논문의 대부분에서 지역 사회와 관련된 내용들이 포함되어 있었다. 다만 Gamma 수치가 네 번째로 높은 논문의 제목이나 내용이 다른 논문들과의 연관성이 상당히 떨어지는 것으로 나타나 이 주제에 대한 명명을 혼란스럽게 하였다. 본 연구에 참여한 전문가 4인 모두 4번째 논문을 본 주제와 관련이 떨어진다는 판단을 하였고 주제에 대한 명명에 이 논문의 내용을 반영하지 않기로 결정을 하였다. 따라서 본 주제는 "Sport in communities"로 명명 되었다.

Topic 9에서는 'fan', 'team', 'factor' 등의 단어가 높은 출현빈도를 보였다. Gamma 수치가 높은 다섯 개의 논문에서는 관람 스포츠의 소비자인 팬들의 행동에 대한 내용이 주를 이루고 있었다. 하지만 Gamma 수치가 세 번째로 높은 논문의 경우에는 관람 스포츠의 소비자가 아닌 참여 스포츠 소비자의 행동에 관련된 논문이어서 관람 스포츠 소비자로 한정지어 주제를 명명하기는 애매한 면이 존재하였다. 추가적으로 Gamma 수치가 6위에서 10위에 있는 논문들의 제목을 점검한 결과 관람스포츠 팬의 내용과 참여스포츠 소비자의 내용이 혼재한 것을 확인하였다. 따라서 본 주제는 "Sport consumers"로 명명하였다.

Topic 10은 'athlet', 'elit', 'perform', 'coach' 등의 단어들이 출현 빈도가 높은 단어로 선정되었다. Gamma 수치가 높게 나타난 상위 다섯 개의 논문들의 주요 주제 또

한 엘리트 스포츠와 관련된 제목들이었다. 또한 두 번째 논문을 제외한 네 가지 논문의 내용이 주로 높은 경기 수행력과 경쟁력 확보에 관한 내용으로 이루어진 것을 확인할 수 있다. Gamma 수치가 높은 상위 다섯 개의 논문 중 두 번째 논문의 제목이 다른 네 개의 논문과는 유사성을 보이고 있지 않아 논문의 내용을 확인하였다. 두 번째 논문의 내용은 스포츠 용품과 관련된 국제 무역과 관련해서 미국, 유럽 및 신흥국가들의 전문성을 다룬 논문이었다. 본 연구에 참여한 4인의 전문가 모두 본 연구의 주제가 전문적인 스포츠와 관련된 장비, 스포츠 상품에 관한 내용도 포함하고 있어서 함께 분류된 논문들과 전혀 다른 주제는 아니라는 것에 의견이 모아졌다. 이에 따라 10번째 주제는 "Elite sports"로 명명 되었다.

논 의

본 연구에서는 스포츠 경영 전문 학술지 중에서 세계적으로 높은 인용지수를 유지하고 있으며 대표 학술지로 인정받고 있는 ESMQ의 연구 주제를 분석 하였다. 분석에 있어서 Pitts et al.(2014)이 연구한 ESMQ에 대한 주제분석을 참조하였는데 본 연구는 그들의 연구와 방법 및 결과 면에서 몇 가지 다른 점이 존재한다.

먼저 ESMQ에 사용된 주제들의 구조적인 분석에 대한 논의를 하고자 한다. 논문이 발행된 년도를 기준으로 각각의 주제들이 몇 편씩 발행되었는지 확인하였는데 2014년, 2016년과 2017년 총 3개년을 제외 하고는 모든 주제들이 빠짐없이 연구되지는 않았던 것으로 나타났다. 3개년에는 모든 주제가 연구되었지만 2014년에는 Sport consumers와 관련된 논문이 6편(23.1%)으로 가장 많이 발행되었고 2016년에는 Cause-related marketing이 7편(21.9%), European sport leagues가 6편(18.9%) 씩 발행된 것으로 나타나 연구자들의 관심을 더 많이 받은 주제인 것으로 보인다.

발행된 연도별로 특별히 많은 논문편수가 분배되지는 않았지만 연구자들에게 꾸준히 다뤄져온 주제도 존재한다. 3번째 주제로 명명된 Factors affecting the results of the competition, 4번째 주제로 명명된 Managing sport organization, 8번째 주제로 명명된 Sport in

communities와 9번째 주제로 명명된 Sport consumers 총 네 개의 주제인데 ESMQ의 저자들은 이 네 개의 주제에 대한 연구를 상당히 지속적으로 해 온 것으로 나타났다. 이러한 현상은 북미를 거점으로 하고 있는 Journal of Sport Management와 약간의 차이를 보이는 것으로 나타났다. JSM의 경우에는 스포츠 조직에 대한 연구와 스포츠 관람객(i.e., Fan)을 대상으로 한 연구는 꾸준히 제시되고 있지만 Factors affecting the results of the competition과 Sport in communities는 꾸준한 관심을 보이는 주제는 아닌 것으로 나타났다.

Pitts et al.(2014)의 연구에서 분류기준으로 사용되었던 주제들의 상대수가 본 연구에서는 주요 연구 주제로 추출되지 않은 것으로 나타났다. 예컨대 Pitts et al.(2014)에 따르면 EJSM에서 단 한 편의 논문이 할당되어 전체 0.1%에 지나지 않던 연구 주제였으나 ESMQ에서는 20편의 논문이 할당되어 전체 9.0%로 연구자들의 관심이 상승한 것으로 나타났던 Governance가 본 연구에서는 주요 주제로 추출되지 않다. 뿐만 아니라 다섯 번째로 높은 관심을 받았던 Sport management education을 포함한 이하 순위의 주제들(i.e., Ethics in sport management, Sport law, Finance in sport, Field experience) 역시 마찬가지이다. 사실 이러한 주제는 전체 연구 주제들 중에서 상위 세 번째 순위로 나타났던 주제들과 비교했을 때 현저히 적은 논문 편수가 할당되어 있었다. 때문에 상기의 주제들이 본 연구에서 주요 연구 주제로 추출되지 않은 것은 어쩌면 당연한 결과라고 볼 수 있다. Pitts et al.(2014)의 연구에서 Sport management education, Ethics in sport management, Sport law, Finance in sport, Field experience 등이 독립적인 주제로 제시되었던 것은 단지 주제 분류 전에 저자들이 이 주제를 NASPE-NASSM SMPS를 바탕으로 만들어 놓았기 때문이다.

하지만 Pitts et al.(2014)의 연구에서 연구자들에게 많은 관심을 받은 것으로 나타났던 Management & organizational skills in sport 주제 영역은 주제 분석에 다른 분류 체계를 적용하여 분석하였음에도 불구하고 주요 연구 주제로 추출되었다. 본 연구에서 추출된 주제 중에 Management & organizational skills in sport 주제 영역으로 볼 수 있는 Managing sport organization

(n=43, 16.2%)과 Strategic management(n=23, 8.7%)를 포함시켜서 비교한다면 Pitts et al.(2014)이 보고한 27.7%(n=85)와 유사한 수준인 24.9%(n=66)로 나타났다. 이와는 다르게 Sport marketing 영역은 전체 논문에서 차지하는 비중이 Pitts et al.(2014)의 연구 결과와 비교 할 때 큰 폭으로 상승한 것으로 나타났다. Sport marketing과 유사한 주제 들이 포함된 Sport economics 주제는 Pitts et al.(2014)의 연구에서 네 번째로 많은 논문이 할당되었던 주제다. 그들에 의하면 전체 ESMQ 논문 중에서 9.5%(n=21)의 논문이 Sport economics에 할당되었는데 본 연구에서도 비슷한 수준의 결과로(n=21, 7.9%) 나타났다. 이와 함께 Impact of mega sport event 주제 논문이 10.2%(n=27), Cause-related marketing 주제 논문이 8.7%(n=23), Sport economics 주제 논문이 7.9%(n=21) 그리고 Sport consumers 주제 논문이 13.6%(n=36)로 나타나 전체 40.4%가 Sport marketing과 관련된 논문이었다. 이는 Pitts et al.(2014)이 보고한 16.0%(n=49)보다 24.4%가 향상된 수치로 함께 추출된 연구 주제 중에 가장 큰 양적 성장이 있던 주제영역으로 볼 수 있다.

상술한 Management & organizational skills in sport 주제영역과 Sport marketing 주제영역은 최근에 보고된 JSM에 대한 주제 분석 결과와 비슷 경향을 보인다(Kwon, Baeck & Choi, 2019). Kwon, et al.(2019)의 연구는 본 연구와 동일한 분석 방법으로 JSM의 주제 분석이 이루어 졌다. 그들의 연구에서 JSM의 주요 연구 주제로 도출된 Management & organizational skills in sport 영역에 대한 결과를 보면 Studies on sport organization 주제 논문이 6.8%(n=25), Diversity management 주제 논문이 10.4%(n=38) 그리고 Role conflict 주제 논문이 6.5%(n=24)로 나타났다. 이 세 개의 주제는 전체 분석된 연구의 23.7%로 선행연구(Pitts & Pedersen, 2005)에서 Management & organizational skills in sport 영역의 연구가 38%(n=89)인 것과 비교할 때 상대적으로 적은 비율이라고 볼 수 있다. 하지만 Sport marketing 주제영역에서는 Management & organizational skills in sport 주제영역과 다른 결과가 확인 되었다. 이 주제영역에는 Sport event & tourism 주제 논문이 7.9%(n=29), Sponsorship

and endorsement 주제 논문이 10.1%(n=37), Consumer behavior 주제 논문 10.9%(n=40), 그리고 Sport service and satisfaction 주제 논문 13.4%(n=49)가 포함 되어 전체 분석된 연구의 42.2%를 차지하는 주제 영역이다. 이는 Pitts와 Pedersen (2005)의 연구에서 Sport marketing 주제영역에 18%(n=41)와 Sport business in the social context 주제영역에 10%(n=24)의 논문이 할당 되어 28%의 비율을 나타낸 것과 비교할 때 괄목할 만한 양적 성장이 있는 주제영역이라고 볼 수 있다. 결과로 미루어 볼 때 ESMQ와 JSM 공통적으로 Management and organizational skills in sport에 대한 연구자들의 관심은 줄고 Sport marketing에 대한 연구자들의 관심이 큰 폭으로 증가되고 있다고 볼 수 있다.

이밖에도 ESMQ와 JSM의 연구 주제분석 결과 중에는 ESMQ의 Elite sports(n=20, 7.5%)와 JSM의 Professional sports(n=43, 11.7%)와 같이 비슷한 주제들이 있다. 이는 두 개의 학술지가 스포츠 경영 전문지라는 공통분모를 가지고 있기 때문으로 여겨진다. 물론 ESMQ에 대한 본 연구결과와 JSM(Kwon, et al., 2019)의 주제분석결과에 상이한 부분도 존재한다. 가장 눈에 띄게 다른 결과는 Governance 주제영역에 관한 부분이다. ESMQ에서는 주제영역으로 추출조차 되지 않았지만 JSM에서는 Pitts와 Pedersen(2005)의 연구 결과(n=12, 5%)보다 상당히 많은 양적 성장(n=50, 13.6%)을 나타내며 독립된 주제로 추출되었다. Role conflict(n=24, 6.5%) 역시 JSM에서만 추출된 주제이고 Factors affecting the results of the competition(n=19, 7.2%)과 European sport leagues(n=27, 10.2%)는 ESMQ에서만 추출된 주제이다. 이와 같은 각각의 독특한 주제들은 ESMQ와 JSM의 특색으로 판단된다.

이렇게 선행연구 결과와 본 연구결과와의 유사점 그리고 다른점 들을 제시하였는데, 본 연구결과에 따라 제시된 주제들은 Pitts et al.(2014)의 연구에서 나타난 주제들과 다른 의미를 갖는다. Pitts et al.(2014)등의 연구를 통해서 제시된 주제분석 결과는 NASPE-NASSM SMPS에 해당하는 주제들이 얼마만큼 연구되고 있는지 파악된 것이라고 볼 수 있으며, 본 연구는 실제로 어떠한 주제들이 연구되었었는지에 대한 경험적인 분석결과가 도출되

었다. 그렇다보니 Pitts et al.(2014)등의 연구결과에서 논문이 4편(1.3% - Finance in sport) 혹은 2편(0.7% - Field experience)이 할당된 주제들도 다른 주제들과 나란히 하나의 주제 영역으로 존재하던 것과 같은 현상이 없다. 본 연구에서 제시된 주제들은 주제들 간에 논문편수가 차이는 있지만 최소 19편의 논문이 한 주제(7.2% - Factors affecting the results of the competition)의 영역으로 분류된 것으로 나타났다. 물론 43편의 논문이 한 주제 영역으로 분류되어(16.2% - Managing sport organization) 다른 주제보다 높은 비율을 보이기도 한다. 이것은 해당 주제에 대해 연구자들의 관심의 차이이기 때문에 주제별로 나타나는 비율의 차이는 당연한 결과라고 볼 수 있다. 뿐만 아니라 추출된 각 주제들은 10.0% 내외의 비율을 가지고 있는 것으로 나타나서 연구자들의 관심 주제로 대표성을 갖기에 충분하다고 볼 수 있다.

본 연구는 연구의 대상으로 ESMQ에 출판된 논문들을 사용하였고 이 논문들은 아무래도 출판의 지리적 기원인 유럽지역의 스포츠산업과 현상을 반영하는 연구들이 많이 포함되었다. 따라서 본 연구에서 나타난 연구의 트렌드나 주제에 대한 관심을 우리 국내 학자들도 같은 정도로 갖아야 한다는 주장은 딱히 큰 의의가 있다고 판단되지는 않는다. 다만 본 연구에 대한 주제분석의 내용이 국내에서 이루어지고 있는 스포츠경영과 어떤 주제적 유사성 혹은 주제적 차이점이 있는지는 아직 정확하게 비교하기가 어렵다. 아직 국내의 스포츠경영 관련 논문들을 토픽모델링을 사용하여 분석한 연구가 없다보니 본 연구의 결과와 비교해서 설명될 사전연구가 존재하지 않는다. 따라서 국내 스포츠경영에 대한 주제분석의 연구에 있어서도 기존의 주제분석 방식을 벗어나 텍스트마이닝의 기술이 적용이 된 주제분석이 이루어진다면 기존에 이루어진 북미대륙의 JSM에 대한 토픽모델링이나 ESMQ에 대한 토픽모델링의 결과가 직접적으로 같이 비교될 수 있을 것이다.

결론 및 제언

본 연구에서는 2009년부터 2018년 사이에 ESMQ를 통해 발간된 논문들에서 연구자들에게 중요하게 다뤄지

는 연구 주제들이 무엇인지를 확인하고자 하였다. 이를 위해 토픽모델링 기법을 사용하여 주제분석을 수행하였는데 기존에 수행되었던 연구와 다른 구조의 결과가 제시되었다. 이 같은 결과는 분석방법의 차이에서 기인한 것으로 볼 수 있다. 선행연구에서는 NASPE-NASSM SMPS를 바탕으로 설정한 주제에 각 논문들을 할당하는 방식으로 분석하였고 본 연구는 맥락과 관련된 단서들을 이용하여 해석가능성이 높은 주제를 추출해주는 특성을 가진 LDA 확률모델 알고리즘의 토픽모델링 기법이 적용되었다. 즉, 선행연구에서는 사전에 정의된 범주에 의해 분류되었고 본 연구에서는 토픽의 숫자만 정해서 분류하였다.

바로 이점이 선행연구와 본 연구의 가장 다른 점이다. 선행연구가 적용한 분류방법은 ESMQ에 NASPE-NASSM SMPS의 주제들이 몇 편씩 발행되었는지를 확인한 것이지 ESMQ에서 실제로 어떠한 연구주제들이 다뤄지고 있는지 분석되었다고 보기 어렵다. 뿐만 아니라 선행연구에서 적용되었던 분류방법은 연구자의 통찰력에 의존한 분석으로 타당성 확보에 제한점을 가질 수밖에 없다. 물론 이러한 제한점을 최소화하기 위하여 코드 간에 사전에 정의된 코드 혹은 의미 범주를 설정하지만 타 연구자에 의한 재생가능성의 문제점을 완전히 해결하지는 못한다.

하지만 토픽모델링을 통한 주제분석도 객관성에 대한 상대적 약점을 지니고 있다. 자동으로 분류되어 추출된 단어들과 10개의 묶음으로 분류된 논문들 그리고 그 논문들이 주제와 얼마만큼 깊은 관계가 있는지 나타내 주는 Gamma 수치를 바탕으로 주제들을 명명하는 과정에서 연구자의 통찰력이 요구되었다. 특히 같은 주제로 묶인 논문 중에서 여타의 논문들과 개연성이 낮다고 판단되는 논문에 대한 해석은 전문가들의 식견에 의존하였다. 이처럼 LDA 알고리즘이 갖는 특성이 분석 대상인 논문들에 대한 사전 지식이 많이 요구되지 않고도 주제분석이 가능하다는 것이지만 때때로 논문이 갖는 의미를 이해하기 위한 추가적인 노력이 필요하였다. 사실 이 과정에서 요구되는 전문가의 식견은 토픽모델링을 통한 주제분석 연구에서 공통적으로 발견되고 있다.

향후 연구에서는 좀 더 명확한 주제 추출을 위하여 토픽 수를 다양하게 설정할 필요가 있다. 본 연구에서는 분

석의 용이성과 일반적으로 주제분석을 하는 연구에서 선정하는 토픽의 수를 고려하여 10개의 토픽으로 주제분석을 수행하였다. 즉, 한가지의 렌즈만을 사용하여 ESMQ에서 발행된 논문들을 들여다 본 것인데, 향후 연구에서는 토픽의 수를 두 가지 이상으로 설정하여 두 개 이상의 렌즈로 들여다본다면 전문가에 의해 명명된 주제에 타당성이 확보될 것이라 판단된다.

본 연구에서는 명확한 주제 추출을 위하여 자료 전처리 과정에 많은 시간이 투자되었다. 265개의 PDF파일을 Text File로 변환한 후 참고문헌과 반복적으로 삽입되어 있는 학술지명, 저자명 등을 수작업으로 삭제하였다. 이는 순수하게 논문만을 대상으로 분석하기 위한 목적이었다. 하지만 LDA는 단순히 단어의 빈도만으로 주제가 추출되는 것이 아니고 문서 내에서 어떠한 단어와 함께 사용되었는지가 고려되는 알고리즘이다. 따라서 순수하게 논문만을 대상으로 주제를 추출하는 것과 참고문헌뿐만 아니라 문서내의 기타 Text들을 제거 하지 않고 주제를 추출하는 경우를 비교 분석하는 연구 혹은 연구자가 선정한 Stopwords를 적용하여 주제를 추출하는 연구가 진행된다면 향후 LDA를 사용하여 주제 분석을 하는 연구자들에게 명확한 주제추출을 위한 가이드라인을 제공하는 연구가 될 것이다.

참고문헌

- Blei, D. (2012). Probabilistic topic models. *Communications of the association for computing Machinery*, 55(4), 77-84.
- CABS (2018). Academic Journal Guide 2018. Retrieved July 5, 2019, from <https://charteredabs.org/academic-journal-guide-2018>.
- Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In *Advances in Neural Information Processing Systems* (pp. 288-296).
- Ciomaga, B. (2013). Sport management: A bibliometric study on central themes and trends. *European sport management quarterly*, 13(5), 557-578.
- Clarivate Analytics (2019). <https://clarivate.com/>
- DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities

- between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding. *Poetics*, 41(6), 570-606.
- ESMQ (2019). Aims and scope. Retrieved August 14, 2019 from <https://www.tandfonline.com/action/journalInformation?show=aimsScope&journalCode=resm20>.
- Kim, A. R., & Cho, W. J. (2013). A Comparative Study of Sport Management Research Trends: Focused on KJSM and JSM. *Korean Journal of Sport Management*, 18(6), 17-28.
- Kwon, H. H., Pyun, D. Y., & Lee, C. W. (2008). What is on the Journal of Sport Management? The most popular topics and contexts in the last decade. *Journal of Leisure and Recreation Studies*, 32(4), 231-241.
- Kwon, H. H., Baeck, J. H., & Choi, M. H. (2019). Research topic analysis of the Journal of Sport Management using Topic Modeling. *Korean Journal of Sport Management*, 24(4), 787-33.
- Parkhouse, B. L. (2005). *The management of sport: Its foundation and application* (4th ed.). Singapore: McGraw-Hill.
- Parks, J. B., Zanger, B. R. K., & Quarterman, J. (1998). *Contemporary Sport Management*. Champaign, IL: Human Kinetics
- Pitts, B., Danylchuk, K., & Quarterman, J. (2014). A content analysis of the European Sport Management Quarterly and its predecessor the European Journal for Sport Management: 1984-2012. *Sport Management International Journal*, 10(2), 45-72.
- Pitts, B., & Pedersen, P. M. (2005). Examining the body of knowledge in sport management: A content analysis of the Journal of Sport Management. *Sport Management and Related Topics Journal*, 2(1), 33-52.
- Shilbury, D. (2011). A bibliometric analysis of four sport management journals. *Sport Management Review*, 14(4), 434-452.
- Shilbury, D., & Rentschler, R. (2007). Assessing sport management journals: A multi-dimensional examination, *Sport Management Review*, 10, 31-44.
- Slack, T., & Parent, M. M. (2006). *Understanding sport organization: The application of organization theory* (2nd ed.). Champaign, IL: Human Kinetics.
- Sport Management Program Review Council (2000). Sport management program standards and review protocol. national association for sport and physical education and northern American society for sport management: Reston, VA.
- Van Doren, C. S., & Heit, M. J. (1973). Where it's at: A content analysis and appraisal of the Journal of Leisure Research. *Journal of Leisure Research*, 5(1), 67-73.

European Sport Management Quarterly의 연구 주제 분석: Latent Dirichlet Allocation을 사용한 Topic Modeling

최미화 · 권형일 · 백주혜(중앙대학교, 박사과정·교수·박사과정), 편도영(러프버러대학교, 교수)

【목적】 본 연구는 2009년부터 2018년까지 European Sport Management Quarterly(ESMQ)에서 발간된 논문들의 주제를 분석하는데 목적이 있다. ESQM에 대한 주제분석이 전무했던 것은 아니지만 실제 어떠한 연구가 이루어지고 있는지에 대한 분석이라기보다는 NASPE-NASSM SMPS를 바탕으로 설정한 주제에 각 논문들을 할당하는 방식으로 주제분석이 이루어졌다. **【방법】** 본 연구에서는 단어들이 어떤 단어들과 함께 사용되고 있는가를 바탕으로 주제를 추출하는 Latent Dirichlet Allocation(LDA)기반의 토픽모델링(topic modeling)을 이용하여 분석하였다. 이를 위해 PDF파일로 확보한 265개의 논문을 ANSI 형식의 Text File로 변환 시켰으며 분석은 R 프로그램의 스크립트를 이용하였고 topic 수는 10개로 설정하였다. **【결과】** 10개로 분류된 논문들을 Gamma 수치가 높은 순으로 재정렬하여 각각의 주제를 명명하는데 기준을 삼았는데 각각의 주제는 (1) Impact of mega sport event, (2) Cause-related marketing, (3) Factors affecting the results of the competition, (4) Managing sport organization, (5) European sport leagues, (6) Strategic management, (7) Sport economics, (8) Sport in communities, (9) Sport consumers, (10) Elite sports로 명명되었다. 선행연구의 분류기준과 본 연구의 분류기준이 다르기 때문에 두 개의 연구를 직접적으로 비교하는 것은 불가능하다. 하지만 분류기준이 다름에도 불구하고 Sport marketing 주제 영역은 여타의 주제들에 비해 가장 큰 양적 성장이 있던 연구주제인 것으로 나타났다. **【결론】** 객관적이고 명확한 주제분석을 위하여 LDA 확률모델 알고리즘의 토픽모델링 기법이 적용되었으나 연구자의 통찰력은 여전히 필요하였다. 이에 향후 연구에서는 다양한 토픽 수로 분석하여 비교하거나 데이터의 전처리 여부에 따른 주제 추출 결과 비교를 통해 주제분석의 객관성 확보에 근거가 되기를 기대한다.

주요어: 주제분석, 텍스트 마이닝, 토픽 모델링