

잠재디리클레할당을 이용한 한국학술지인용색인의 풍력에너지 문헌검토

김현구^{1)*} · 이제현²⁾ · 오명찬³⁾

Review of Wind Energy Publications in Korea Citation Index using Latent Dirichlet Allocation

Hyun-Goo Kim^{1)*} · Jehyun Lee²⁾ · Myeongchan Oh³⁾

Received 27 August 2020 Accepted 9 November 2020 Published online 26 November 2020

ABSTRACT The research topics of more than 1,900 wind energy papers registered in the Korean Journal Citation Index (KCI) were modeled into 25 topics using latent directory allocation (LDA), and their consistency was cross-validated through principal component analysis (PCA) of the document word matrix. Key research topics in the wind energy field were identified as “offshore, wind farm,” “blade, design,” “generator, voltage, control,” “dynamic, load, noise,” and “performance test.” As a new method to determine the similarity between research topics in journals, a systematic evaluation method was proposed to analyze the correlation between topics by constructing a journal-topic matrix (JTM) and clustering them based on topic similarity between journals. By evaluating 24 journals that published more than 20 wind energy papers, it was confirmed that they were classified into meaningful clusters of mechanical engineering, electrical engineering, marine engineering, and renewable energy. It is expected that the proposed systematic method can be applied to the evaluation of the specificity of subsequent journals.

Key words Wind Energy(풍력에너지), Korea Citation Index(KCI, 한국학술지인용색인), Text Mining(텍스트마이닝), Topic Modeling(토픽모델링), Latent Dirichlet Allocation(LDA, 잠재디리클레할당), Document-Term Matrix(DTM; 문서 단어행렬), Journal-Topic Matrix(JTM; 학술지토픽행렬)

1. 서론

한국신·재생에너지학회는 국내 신재생에너지 보급과 기술개발을 지원하기 위해 2004년에 창립되었으며, 2005년

에는 신재생에너지를 종합적으로 다루는 학술지인 ‘신재생 에너지(New & Renewable Energy)’를 창간하였다. 2010년에는 학술연구재단의 등재후보학술지로 선정되었고 2015년에는 등재학술지로 승격되었다. 한국학술지인용색인(KCI; Korea Citation Index)의 5년간 인용지수(impact factor)는 0.43으로 공학분야 학술지 중 중상위권(76/233)에 해당한다.

최근 텍스트마이닝을 이용함으로써 기존의 수동적 문헌 검토에서는 불가능하였던 수 백편 이상의 논문을 단시간 내에 분석할 수 있게 되었다. 즉, 패턴매칭 기법을 이용하여 문헌으로부터 특정한 정보를 추출하거나 유의미한 정보

1) Principal Research/Center Chief, New & Renewable Energy Resource Map Laboratory, Korea Institute of Energy Research

2) Senior Researcher, Platform Technology Laboratory, Korea Institute of Energy Research

3) Postdoctoral Researcher, New & Renewable Energy Resource Map Laboratory, Korea Institute of Energy Research

*Corresponding author: hyungoo@kier.re.kr

Tel: +82-42-860-3376 Fax: +82-42-860-3642

를 찾아내고 이를 가시화할 수 있다. 또한 패턴이나 특성을 이용하여 분류를 하거나 군집분류를 할 수 있다. 이러한 기법을 이용하면 문헌의 핵심정보를 체계적으로 요약할 수 있다^[1].

텍스트마이닝 기법 중 문헌 집합의 추상적인 주제(topic)를 발견하기 위한 통계적 모델 중 하나로, 텍스트 본문의 숨겨진 의미구조를 발견하기 위해 사용되는 기법을 토픽모델링이라고 한다. 최근 학술논문의 주제를 찾아내기 위해 토픽모델링을 사용하는 사례가 많아지고 있다.

국내에서는 Shin et al.(2015)이 언론기사를 이용하여 토픽모델링을 통해 신재생에너지 동향을 파악하였으며^[2], Lee et al.(2017)은 특허 및 논문을 이용하여 태양광 기술 동향을 파악하였다^[3]. Yoo et al.(2019)은 한국기후변화학회지 논문으로부터 기후변화 연구동향을 검토하였으며^[4], Kim(2020)은 한국태양에너지학회 논문집의 풍력분야 논문에 대해^[5], Kim et al.(2020a)은 한국풍력에너지학회 풍력에너지저널의 논문주제를 토픽모델링하고 전문가의 수동적 분류와 비교하였다^[6]. 해외에서는 Yang et al.(2018)이 텍스트마이닝으로 태양일사와 태양광 발전예보의 연구동향을 분석한 바 있으며^[7], Xu et al.(2020)은 문헌검토를 통해 재생에너지의 연구동향을 가시적으로 지도화 하였다^[8].

본 연구에서는 한국신·재생에너지학회의 학술지인 ‘신재생에너지’에 게재된 풍력분야의 논문을 포함하여, 한국학술지인용색인에 등재된 국내 학술지에 게재된 풍력관련 논문 전체를 대상으로 토픽모델링을 수행하여 풍력분야 연구의 주제와 동향을 파악하였다. 또한 풍력분야를 다루고 있는 학술지간 논문주제의 유사성을 파악하기 위한 새로운 평가법으로 토픽모델링과 연계한 상관성 분석 및 군집분석을 제안하였다.

2. 연구자료 및 방법

2.1 특성추출

한국학술지인용색인(KCI) 데이터베이스에 등재된 전체 학술지로부터 풍력 관련 논문을 모두 수집하고, 논문의 서지사항인 제목과 키워드로부터 문헌정보의 특성추출(feature extraction)을 하였다. 이때 텍스트마이닝의 효율성을 고

려하여 모든 텍스트는 영문으로 진행하였다. 텍스트에 사용된 동사와 명사의 단어 빈도수를 계산하여 문서단어행렬(DTM; document-term matrix)을 구성하였으며, 불용어(stopword)를 제외하고 어간추출(stemming) 및 단수(singular) 처리를 하는 등 전처리 과정을 거쳤다. 또한 학술논문에서 빈번하게 사용되나 논문의 주제를 특정하지는 못하는 ‘analysis’, ‘calculate’, ‘develop’ 등의 통상적 단어는 분석에서 제외하였다.

DTM의 통계학적인 설명력을 분석하기 위하여 주성분분석(PCA; principal component analysis)을 수행하였으며, 이를 토픽모델링에서 도출된 연구주제와의 정합성 검증에 활용하였다.

2.2 토픽모델링

최근 문헌정보학 연구동향 파악^[9], 특허의 기능적 분류^[10], 학술지 논문의 주제분류^[11] 등에 가장 많이 사용되고 있는 토픽모델링 방법인 잠재디리클레할당(LDA; latent Dirichlet allocation)은 특정 단어가 특정 토픽에 존재할 확률과 문서에 특정 토픽이 존재할 확률간의 결합확률을 디리클레분포(Dirichlet distribution)로 표현하고, 이로부터 토픽을 추출하는 확률기반 알고리즘이다^[12]. Kim et al.(2020b)은 학술논문의 주제분류시 토픽모델링 방법으로 많이 사용되는 잠재의미분석(LSA; latent semantic analysis)에 비해 LDA가 월등히 우수함을 확인한 바 있다^[13].

Fig. 1은 D 개의 문서(θ)와 K 개의 토픽(ϕ)이 단어 w 와 결합확률 $p(\phi, \theta, z, w)$ 로 연결되는 LDA 분석모형을 보여준다.

$$p(\phi, \theta, z, w) = \prod_{i=1}^K p(\phi_i | \beta) \prod_{d=1}^D p(\theta_d | \alpha) \cdot \left\{ \prod_{n=1}^N p(z_{d,n} | \theta_d) p(w_{d,n} | \phi_{1:K} z_{d,n}) \right\} \quad (1)$$

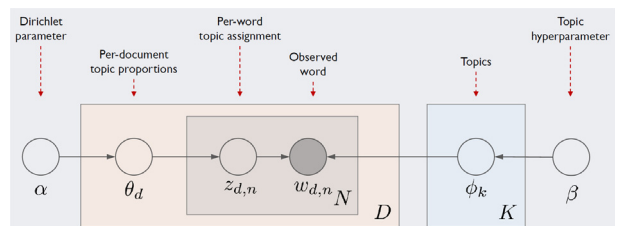


Fig. 1. Topic modeling process of latent Dirichlet allocation

2.3 학술지 유사성 분석

일반적으로 학술지의 주제분야는 상호 독립적이며 특수성을 갖는 것이 바람직하다. 이에 한국연구재단의 학술지 평가기준에서도 “학문분야 특수성의 학술적 가치”를 중요한 정성평가 항목으로 채택하고 있다.

풍력분야에 특화된 학술지로는 한국풍력에너지학회의 풍력에너지저널이 있다. 그러나 풍력은 다양한 세부분야로 구성된 다학제적 공학이므로 각각의 세부분야에 특화된 학술지에 논문을 게재하는 것이 보다 일반적이다. 그런데 한국 학술지인용색인(KCI) 검색결과에 의하면 268개 학술지에 풍력분야 논문이 게재되었으며, 이는 세부주제라고 할지라도 유사한 논문주제를 다루는 학술지가 많음을 시사한다.

본 연구에서는 학술지간 풍력분야 연구주제의 유사성을 판별하기 위한 새로운 방법을 제안하였다. 즉, LDA로 분석한 토픽의 구성단어에 산출된 가중치를 적용하여 논문의 핵심토픽을 선정하고, 이를 바탕으로 학술지토픽행렬(JTM; journal-topic matrix)을 구성하였다. JTM을 수학적으로 표현한다면 학술지의 토픽별 출현빈도를 특성벡터로 구성한 것으로, 학술지간 유사성을 상관도로 추정하였으며 군집분석으로 유사한 학술지를 군집화 하였다. 동일한 방법으로 토픽간의 유사도를 상관계수로 계량할 수 있다.

학술지 논문주제의 체계적인 유사성 판별을 위해 개체를 유사한 것끼리 묶어주는 군집분석을 수행하였다. 유사성의 척도는 코사인(cosine) 거리를 적용하였다^[14]. 계층적 방법과 비계층적 방법으로 각각 응집형계층군집화(AHC; agglomerative hierarchical clustering)와 퍼지 K-평균군집화(fuzzy K-means clustering)를 적용하여 상호 검증하였다. AHC의 개체간 합체방법으로는 비가중쌍집단 평균(unweighted pair-group average)을 이용하였다. 퍼지 K-평균군집화는 군집간의 경계가 모호할 경우 개체간 확률로 정의되는 퍼지계수를 이용하여 K-평균군집화의 단점을 보완한 방법이다.

3. 연구결과

3.1 ‘신재생에너지’ 논문검토

‘신재생에너지’에는 지난 15년간 490여편의 논문이 계

재되었으며, 이 중 풍력분야 논문은 18%인 90편이 게재되었다. 풍력분야 논문의 연구소:대학교:산업체 게재비율은 47:36:17로 연구소의 기여가 월등히 높다. 특히 한국에너지기술연구원은 전체 논문의 25%를 차지한다. 주요 저자기관으로 연구소에서는 한국에너지기술연구원, 항공우주연구원, 기계연구원, 전력연구원 순으로, 대학교에서는 포항공과대학교, 군산대학교, 서울대학교 순으로, 산업체는 유니슨인 것으로 파악되었다. 이들 주요기관 8개의 합산 기여율은 68%이고, 주요저자 8명의 합산 기여율은 58%로, 학술지의 풍력분야 저자폭은 좁은 것으로 판단된다.

Fig. 2는 풍력분야의 논문게재 동향을 도시한 것이며 실선은 풍력분야 논문의 비율을 나타내는데, 풍력분야는 평균 6편/년의 논문이 게재되고 있다.

Fig. 3은 Kim et al.(2020a)이^[6] 풍력분야 논문의 주제분류에 사용한 분류기준을 차용하여 분석한 결과로 9개의 주제분야 중 ‘resource’와 ‘blade’의 비율이 가장 높게 나

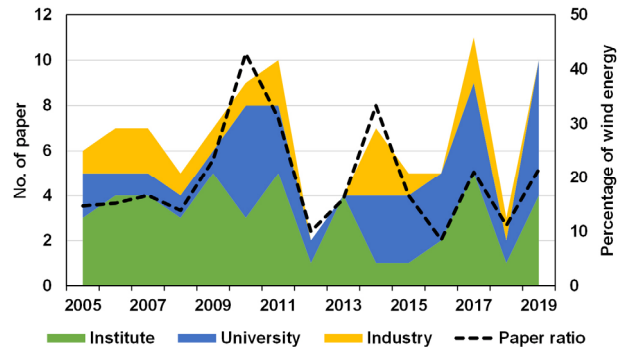


Fig. 2. Publication trend of wind energy articles in New & Renewable Energy

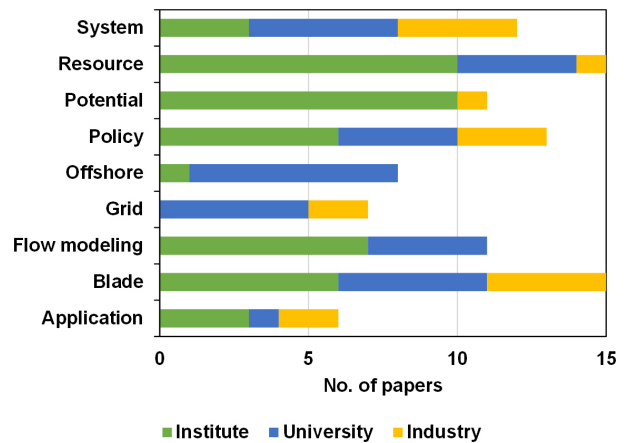


Fig. 3. Classification of wind energy articles by topic

타났다. 기관별 게재특성을 보면 ‘resource’와 ‘potential’은 압도적으로 연구소에서, ‘offshore’와 ‘grid’는 대학교에서 주로 논문을 게재하였다. 산업체는 풍력터빈 제품과 관련된 ‘blade’와 ‘system’ 분야의 참여도가 높은 것으로 나타났다.

3.2 풍력분야 논문검토

한국학술지인용색인(KCI)에 수록된 논문 중 풍력에너지 분야는 2002년부터 2020년까지 268개 학술지에서 총 1,902편이며, Fig. 4와 같이 현재 해마다 100여 편의 풍력 논문이 게재되고 있다. 논문은 ‘풍력’, ‘wind energy’를 키워드로 검색하여 수집하였고 분석과정에서 관련성이 없는 소수의 논문을 배제하였다.

연간 10편 이상 풍력분야 논문이 게재된 학술지는 37종으로, 이들의 게재 편수가 전체의 71%를 차지한다. 풍력분야 논문이 가장 많이 게재된 학술지는 풍력에너지저널(190편), 한국태양에너지학회 논문집(139편), 신재생에너지(90편), 전기학회논문지(86편) 순으로 이들은 전체의 27%를 차지한다.

풍력분야 논문이 게재된 학술지를 살펴보면, 기계공학, 재료공학, 건축공학, 전기공학, 해양수산 등 공학분야뿐만 아니라 경영학, 디자인, 법학, 교육학 등 다학제적인 학문 분야임을 확인할 수 있다.

풍력 논문이 10편 이상 게재된 37종 학술지의 KCI 인용횟수 순위는 환경과학회지(4.3), 한국전자통신학회 논문지(3.6), 한국풍공학회지(3.1), 에너지공학(2.8), 한국태양에너지학회 논문집(2.6) 순이며 신재생에너지(1.1), 풍력에너지저널(0.7), 전기학회논문지(0.7)는 각각 30, 35, 36위로

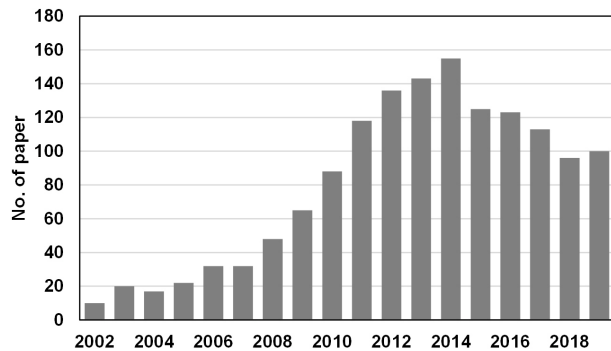


Fig. 4. Publication trend of wind energy articles in KCI database

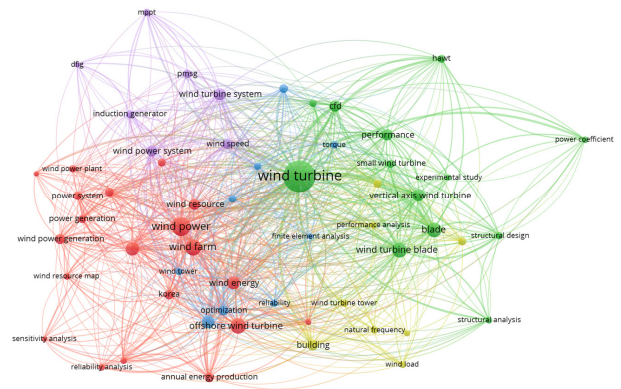


Fig. 5. Bibliometric network map created by VOSviewer

인용지수가 매우 낮았다.

Fig. 5는 서지사항 지도화(bibliometric mapping) 방법인 VOS(visualization of similarity)를 이용하여 키워드의 연관성을 가시화한 것으로, [14] 핵심 키워드인 ‘wind turbine’을 중심으로 ‘wind power’, ‘wind farm’, ‘offshore wind turbine’, ‘blade’ 등의 주요 키워드가 연결된 네트워크를 보여준다.

3.3 토픽모델링

KCI 데이터베이스로부터 풍력분야 논문의 제목과 키워드를 수집하고 정제과정을 거쳐 10회 이상 사용된 450여 단어를 추출하였다. LDA 분석을 통해 25개의 토픽을 도출하였으며 이들의 설명력은 80% 수준이다. 토픽 구성단어는 상위 5개로 한정할 경우 중복단어의 비율이 25%이며, 이는 LSA의 토픽 구성단어 중복율인 50%에 비해 절반 수준이다. [13] 따라서 2배 더 많은 단어로 토픽을 설명함에 따라 토픽을 보다 특징적으로 설명할 수 있는 것으로 판단된다. 참고로 중복 사용된 토픽 구성단어는 ‘offshore’, ‘design’이 5회, ‘wind farm’이 4회이다.

Table 1은 LDA로 도출된 토픽으로, DTM의 주성분분석 결과와도 대응하였다. 표에 나열된 단어는 가중치가 높은 순위에 따라서이며, 제7 주성분까지의 누적 설명력은 71%이다. 각 토픽이 차지하는 비율은 최소 0.8%에서 최대 9.7%의 편차를 갖지만 평균은 $4.0 \pm 2.4\%$ 로 대체로 균일한 분포를 보인다.

제1 주성분에 대응되는 토픽은 ‘offshore, wind farm’을 핵심 구성단어로 조합된 Topic 4, 6, 13, 14, 17, 23으

Table 1. Topic modeling result by latent Dirichlet allocation

Topic	Topic words					PCA	Ratio (%)
Topic 1	Assessment	Resource	Jeju	Onshore	Potential		5.6
Topic 2	Control	Generator	Operation	Pmsg	Induction	PC3	8.4
Topic 3	Design	Vertical	Axis	Blade	Performance	PC2	9.7
Topic 4	Offshore	Floating	Tower	Design	Dynamic	PC1	7.7
Topic 5	Korea	Improvement	New	Industry	Technology		1.5
Topic 6	Farm	Wake	Field	Large	Complex	PC1	2.7
Topic 7	Test	Performance	Small	Tunnel	Force	PC5	5.7
Topic 8	Performance	Nacelle	Lidar	Comparative	Testing	PC5	2.0
Topic 9	Blade	Structural	Composite	Scale	Design	PC2	7.1
Topic 10	Characteristics	Noise	Numerical	Investigation	Aerodynamic	PC4	1.6
Topic 11	Korea	Resource	Map	Area	Peninsula		3.8
Topic 12	Load	Building	Response	Tall	Vibration	PC7	3.2
Topic 13	Farm	Offshore	Capacity	Production	Jeju	PC1	6.3
Topic 14	Offshore	Farm	Case	Local	Southwest	PC1	2.0
Topic 15	Model	Forecasting	Speed	Simulation	Turbulence	PC6	2.8
Topic 16	Generator	Design	Pitch	Reduction	Permanent	PC3	3.9
Topic 17	Offshore	Foundation	Model	Network	File	PC1	3.0
Topic 18	Characteristics	Simulation	Dynamic	Comparison	Various	PC4	1.5
Topic 19	Characteristics	Load	Conditions	Distribution	Gearbox	PC4	1.8
Topic 20	Tower	Monitoring	Frequency	Properties	Natural		2.5
Topic 21	Voltage	Farm	Grid	Connected	Control	PC3	3.1
Topic 22	Control	Speed	Design	Variable	Fuzzy	PC3	6.2
Topic 23	Offshore	Structure	Support	Reliability	Interaction	PC1	5.8
Topic 24	Hybrid	Storage	Generators	Change	Climate		1.2
Topic 25	Blade	Lightning	Composite	Detection	Reliability	PC2	0.8

로 토픽간에는 피어슨(Pearson) 상관계수 0.60~0.94의 높은 상관도를 갖는다. Fig. 6에 도시한 바와 같이 제 1 주성분(PC1)과 직교성을 갖는 제 2 주성분(PC2)에 대응되는 토픽은 ‘blade, design’을 핵심 구성단어로 조합된 Topic 3, 9, 25으로 상관계수는 0.63~0.91로 매우 높다. 그 외 주성분에 해당하는 토픽은 각각 ‘generator, voltage, control’ (PC3), ‘dynamic, load, noise’(PC4), ‘performance test’ (PC5), ‘forecasting’(PC6), ‘building’(PC7)인 것으로 분석되었다.

Fig. 7은 토픽간의 상관행렬을 가시화한 것으로 빨간색은 양의 파란색은 음의 상관도이며 원의 크기는 상관계수의 크기와 비례한다.

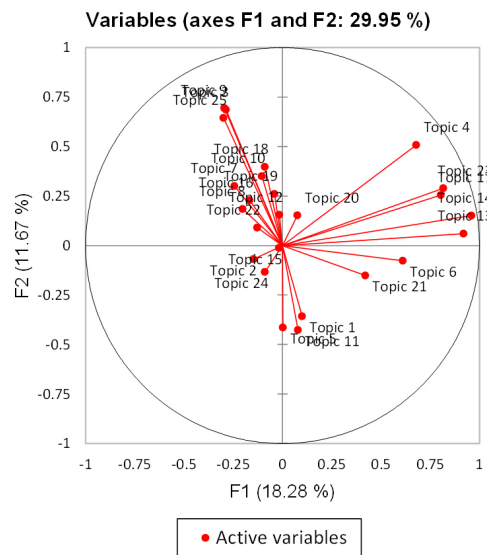


Fig. 6. Principal components on F1 and F2 coordinates

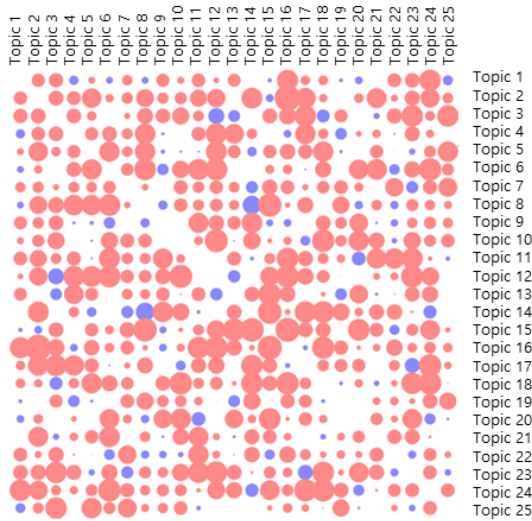


Fig. 7. Correlation matrix between 25 topics

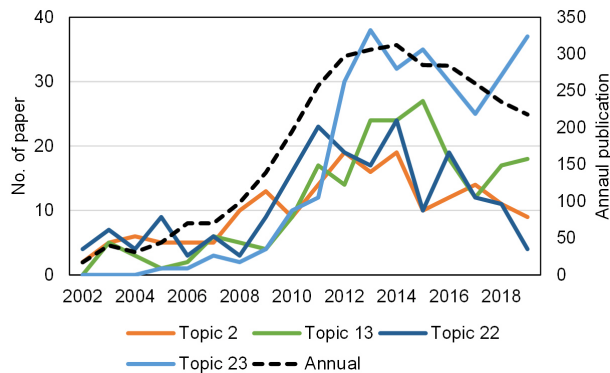


Fig. 8. Publication trends of the most popular topics

3.4 학술지 유사성 분석

유익한 통계분석을 위해 풍력분야 논문이 20편 이상 게재된 24개 학술지에 대해 JTM을 구성하였다. 이들 학술지의 풍력논문 누적비율은 62%이다. JTM으로부터 토픽별 게재 특성을 살펴보면, Topic 2, 13, 22, 23이 6~8개의 학술지에서 공통적으로 집중 게재되었다. 따라서 이들 학술지는 서로 유사성이 높을 것이라고 추정할 수 있다.

Fig. 8은 위 공통 토픽들의 게재추세를 분석한 것으로, 전반적으로 전체 풍력논문의 게재추세(검은색 점선)와 동조되어 나타난다. 그러나 전체 추세를 제외하고 순증가를 보인 토픽은 각각 ‘offshore wind farm’과 ‘offshore structure’를 핵심주제로 다룬 Topic 13과 23이다.

보다 체계적인 유사성 분석을 위해 학술지를 군집분석한 결과, 5개의 군집으로 분류되었다. Fig. 9는 AHC 분류 결과를 보여주지만 퍼지 K-평균군집화도 이와 동일한 결과를 제시하였다. 흥미로운 것은 분류된 5개의 군집, 즉 해양공학분야(7), 환경공학분야(5), 기계공학분야(5), 전기공학분야(5), 기타분야(2)가 유의미한 관련성을 갖는다는 것이다. 풍력에너지는 재생에너지의 한 분야로서 기계공학(풍력터빈, 블레이드), 전기공학(발전기, 전력계통), 해양공학(해상풍력, 하부구조) 등으로 대분류할 수 있으며, 군집분석 결과는 이러한 공학적 특성에도 정확하게 부합된다.



Fig. 9. Dendrogram of the Journals publishing wind energy by AHC

기계공학분야 군집은 대한기계학회논문집, 한국유체기계학회 논문집, 한국항공우주학회지, 한국정밀공학학회지로, 이들은 기계공학 중 유체역학 분야의 대표적 학술지로 ‘blade’가 공통주제이다.

전기공학분야 군집은 전기학회논문지, 조명·전기설비학회논문지, 한국전자통신학회 논문지와 한국산학기술학회 논문지로 분류되었다. 한국산학기술학회논문지에 게재된 풍력논문도 대부분 전기공학이 핵심주제이다.

해양공학분야 군집은 한국해양공학학회지, 한국해양·해양공학학회논문집, 한국마린엔지니어링학회지 등 해양공학분야의 대표적인 학술지와 한국전산구조공학학회논문집, 한국소음진동공학회 논문집, 한국방재학회논문집, 풍력에너지저널 등 그 외의 학술지로 분류되었다. 한편 풍력에너지저널을 문헌검토한 선행연구에 의하면 학술지의 핵심주제가 해상풍력으로 판별되었다.^[6] 따라서 풍력에너지저널이 해양공학분야의 군집으로 분류된 것은 유의미한 연관성이 있다. 한국전산구조공학학회논문집도 해상풍력의 구조해석을 핵심주제로 다루고 있으며, 한국방재학회논문집은 지진 등에 의한 해상구조물 안전성이라는 연결성이 있다.

재생에너지분야의 군집은 풍력을 비롯한 신재생에너지 전반을 다루고 있는 신재생에너지, 한국태양에너지학회 논문집과 한국환경과학회지, 해양환경안전학회지, 한국풍공학회지가 분류되었는데, 이들은 핵심 연구주제가 풍력자원이라는 연결성을 갖는다.

4. 결론 및 제언

본 연구에서는 한국학술지인용색인(KCI)에 등재된 풍력에너지 분야의 1,900여 편 논문을 토픽모델링 하였으며, 다음과 같은 결론을 도출하였다.

- 1) 풍력에너지는 다학제적 공학분야로서, 한국학술지인용색인에 등재된 268개 학술지에 1,900여 편의 논문이 등록되어 있으며 매년 100여 편의 논문이 게재되고 있다. 한국신·재생에너지학회의 학술지인 신재생에너지에도 90여 편의 논문이 게재되었으며 이들의 핵심주제는 ‘wind resource’와 ‘blade’인 것으로 분

석되었다.

- 2) 토픽모델링 방법 중 논문 주제분류에 적합성이 높은 잠재디리클레할당(LDA)을 적용하여 25개의 토픽을 도출하였으며, 문헌단어행렬(DTM)의 주성분분석(PCA)과도 정합성을 갖는 연구주제 분류가 되었음을 확인하였다. 풍력논문의 핵심주제는 ‘offshore, wind farm’, ‘blade, design’, ‘generator, voltage, control’, ‘dynamic, load, noise’, ‘performance test’인 것으로 파악되었다.
- 3) 학술지간 논문주제의 유사성을 판별하기 위해 학술지 토픽행렬(JTM)을 이용한 상관분석 및 군집분석 방법을 제안하였다. 풍력논문이 20편 이상 게재된 24개 학술지에 대한 분석 결과, 기계공학, 전기공학, 해양공학, 재생에너지 등으로 분류가 되며 이러한 분류는 공학적 특성이 반영된 유의미한 결과라고 사료된다.

유사한 논문주제를 다루는 학술지가 많은 경우, 논문심사를 위한 전문가 풀(pool)이 공유되지 못하는 등의 문제로 논문의 질적인 저하뿐 아니라 논문의 산재로 인한 활용성 저하가 우려된다. 따라서 본 연구에서 학술지간 논문주제의 유사성을 판별할 수 있는 체계적인 데이터마이닝 평가법이 새롭게 개발되었다는 것은 중요한 의미가 있다.

후 기

본 논문은 한국에너지기술연구원의 주요사업으로 수행한 결과입니다(C0-2407).

References

- [1] Feng, L., Chian, Y.K., and Lo, S.K., 2017, “Text mining techniques and tools for systematic literature reviews: a systematic literature review”, 2017 24th Asia-Pacific Software Engineering Conference (APSEC), 41-50, doi: 10.1109/APSEC.2017.10.
- [2] Shin, K.S., Choi, H.R., and Lee, H.C., 2015, “Topic model analysis of research trend on renewable energy”,

- J. of the Korea Academic- Industrial Cooperation Society, **16**(9), 6411-6418.
- [3] Lee, J.H., Lee, I.S., Jung, K.S., Chae, B.H., and Lee, J.Y., 2017, "Patents and papers trends of solar-photovoltaic (PV) Technology using LDA algorithm", J. of Digital Convergence, **15**(9), 231-239.
- [4] Yoo, J.H., Jeon, E.C., and Kim, H.N., 2019, "Study of research trends in climate change using text analysis - focusing on journal of climate change research", J. of Climate Change Research, **10**(3), 161-172.
- [5] Kim, H.G., 2020, "Review of trends in wind energy publication in journal of the Korean solar energy society", J. Korean Solar Energy, **40**(4), 1-11.
- [6] Kim, H.G., Ryu, K.W., and Paek, I.S., 2020a, "Topic modeling with a literature review of the journal of wind energy", J. of Wind Energy, **11**(2), 30-36.
- [7] Yang, D., Kleissl, J., Gueymard, C.A., Pedro, H.T.C., and Coimbra, C.F.M., 2018, "History and trends in solar irradiance and PV power forecasting: a preliminary assessment and review using text mining", Solar Energy, **168**, 60-101.
- [8] Xu, W., Guo, L., and Liang, L., 2020, "Mapping the academic landscape of the renewable energy field in electrical and electronic disciplines", Appl. Sci., **10**(8), 2879.
- [9] Park, J.H. and Song, M., 2013, "A study on the research trends in library & information science in Korea using topic modeling", J. of the Korean Society for Information Society, **30**(1), 7-32.
- [10] Cvitanic, T., Lee, B.S., Song, H.I., Fu, K., and Rosen, D., 2016, "LDA vs. LSA: A Comparison of two computational text analysis tools for the functional categorization of patents", Proceedings of the ICCBR 2016 Workshops, Atlanta, USA.
- [11] Cho, K.W., Bae, S.K., and Woo, Y.W., 2017, "Analysis on topic trends and topic modeling of KSHSM journal papers using text mining", The Korean J. of Health Service Management, **11**(4), 213-224.
- [12] Blei, D.M., Ng, A.Y., and Jordan, M.I., 2003, "Latent Dirichlet Allocation", J. Mach. Learn. Res., **3**, 993-1022.
- [13] Kim, H.G., Hwang, J.K., and Hwang, J.S., 2020b, "Topic modeling of journal of the wind engineering institute of Korea using LSA and LDA", J. Wind Eng. Inst. Korea, **24**(3), 113-120.
- [14] Van Eck, N.J., Waltman, L., Dekker, R., and Van den Berg, J., 2010, "A comparison of two techniques for bibliometric mapping: multidimensional scaling and VOS", J. Am. Soc. Inf. Sci. Tec., **61**(12), 2405-2416.