

연구자의 논문 게재 이력을 고려한 저널 결정 요인별 중요도 학습 기반의 저널 추천 방법론[☆]

Development of Journal Recommendation Method Considering Importance of Decision Factors Based on Researchers' Paper Publication History

손 연 빈¹ 장 태 우² 최 예 립³
Yeonbin Son Tai-Woo Chang Yerim Choi

요 약

연구자는 논문을 투고할 저널을 선택하는 과정에서 저널의 수가 방대하다는 점, 고려할 저널 결정 요인이 다양하다는 점에서 어려움을 겪는다. 이러한 어려움을 해소하기 위해 IRA(intelligent research assistant)의 한 종류로 연구자별로 논문 투고에 적합한 저널을 추천해주는 저널 추천 서비스를 활용할 수 있다. 하지만 현재 운영 중인 저널 추천 서비스의 경우 주제 유사도 및 수치적 필터링을 기반으로 저널 추천을 실행하고 있으며, 이 경우 연구자가 논문 데이터를 입력하지 않으면 주제 유사도를 고려할 수 없고, 수치적 필터링 기능도 연구자 스스로 결정 요인별 수치 범위를 명확하게 정하기에 어려움이 있다는 한계점이 존재한다. 따라서 본 논문에서는 연구자의 논문 게재 이력을 이용해 선호도 행렬을 형성하고, 이를 기반으로 저널 별 선호 점수를 고려한 저널 추천 방법론을 제안한다. 연구자는 다수의 저널 결정 요인에 대해 상이한 중요도를 가지고 있는데, 결정 요인 별 선호 민감도를 계산해 중요도를 학습한 뒤 이를 기반으로 모든 저널에 대한 선호 점수를 도출하여 저널을 추천한다는 점에서 의의가 있다. 실제 데이터를 이용하여 저널 추천 실험을 수행했으며 제안 방법론의 우수한 성능을 확인하였다.

☞ 주제어 : 저널 추천, 결정 요인별 중요도, 논문 게재 이력, 머신러닝, 데이터 분석

ABSTRACT

Selecting a proper journal to submit a research paper is a difficult task for researchers since there are many journals and various decision factors to consider during the decision process. For this reason, journal recommendation services are exist as a kind of intelligent research assistant which recommend potential journals. The existing services are executing a recommendation based on topic similarity and numerical filtering. However, it is impossible to calculate topic similarity when a researcher does not input paper data, and difficult to input clear numerical values for researchers. Therefore, the journal recommendation method which consider the importance of decision factors is proposed by constructing the preference matrix based on the paper publication history of a researcher. The proposed method was evaluated by using the actual publication history of researchers. The experiment results showed that the proposed method outperformed the compared methods.

☞ keyword : journal recommendation, importance by decision factors, paper publication history, machine learning, data analysis

1. 서 론

연구자는 연구 발표를 목적으로 논문을 투고할 저널을 선택하는데, 저널의 수가 많고 연구자마다 환경적 제한이 있어 고려해야하는 요소가 다양하므로 저널을 결정하는 과정에서 많은 어려움을 겪는다. 2019년 현재, SCI급 저널의 수가 1만 3천여 개에 달할 정도로 저널의 양은 방대하기 때문에[1], 연구자가 연구 논문을 투고할 저널을 결

우수 논문 추천에 따라 확장 및 수정된 논문임.

1 Department of Industrial and Management Engineering, Kyonggi University, Suwon, Republic of Korea

2 Professor, Department of Industrial and Management Engineering, Kyonggi University, Suwon, Republic of Korea

3 Assistant Professor, Department of Industrial and Management Engineering, Kyonggi University, Suwon, Republic of Korea

* Corresponding author: (yrchoi@kgu.ac.kr)

[Received 17 April 2019, Reviewed 25 April 2019(R2 9 June 2019), Accepted 3 July 2019]

☆ This work was supported by Kyonggi University's Graduate Research Assistantship 2018.

☆ 본 논문은 2018년도 한국인터넷정보학회 추계학술발표대회

정하기 위해 자신의 연구 분야에 속하는 모든 저널의 정보를 확인하는 것은 불가능한 일이다[2]. 또한, 여러 연구자가 같은 분야의 연구를 진행한다 하더라도 연구자마다 상황이 다르기 때문에 중요하게 생각하는 결정 요인이 상이하다. 예를 들면, 연구 실적을 채우기 위해 논문을 저널에 투고하는 연구자의 경우 논문 투고 프로세스의 속도가 중요하며, 연구적 성과를 중요시하는 연구자의 경우 저널의 IF(impact factor)를 중요시 할 것이다. 하지만 연구자는 실제로 자신이 어떤 결정 요인을 어느 정도로 중요하게 생각하는지에 대해 확실하게 알지 못하기 때문에 다양한 결정 요인을 명확하게 고려하기 어렵다.

이러한 연구 논문 작성 및 투고 등의 연구 과정에서 발생하는 어려움을 해소하고자 다수 종류의 IRA(intelligent research assistant)가 있으며 많은 연구자가 이용하고 있다. IRA 별로 목적은 다양한데, 첫 번째로 여러 분야의 연구자가 함께 연구 프로젝트를 진행할 수 있는 플랫폼이 존재한다[3]. 또한 논문의 주제를 기반으로 비슷한 논문을 검색하여 제시하는 플랫폼이 있다[4]. 연구 논문의 효과적인 의미 전달을 위해 모호한 표현과 어휘, 잘못된 문법 등에 대해 논문 교정을 진행하는 서비스가 있는데, 이러

한 논문 교정 서비스를 제공하는 회사가 이용할 수 있는 주제 기반의 논문 교정자 매칭 방법론이 발표된 바 있다[5]. 마지막으로, 연구자의 논문과 필터링 기능을 이용해 저널을 추천해주는 서비스도 존재한다[6,7].

본 논문에서는 그 중에서도 저널 추천 서비스에 주목한다. 일반적인 저널 추천 서비스의 경우, 논문의 내용을 입력하고 추가적으로 결정 요인에 대한 수치적 필터링을 진행한다. 그림 1은 이에 대한 서비스의 예시이다.

논문의 내용이라 함은 제목, 초록, 키워드 등을 의미하며, 텍스트 분석을 통한 주제 유사도 기반 추천이 이루어진다. 하지만 일반적으로 연구자는 연구에 대한 아이디어를 기반으로 어떤 성질을 가지는 저널에 투고할지 선택한 후 논문을 작성하기 때문에 연구 논문을 완벽하게 작성하기 전 서비스를 이용하는 경우가 존재하며, 이 경우 본 서비스를 이용하기에 어려움이 있다. 두 번째로 수치적 필터링의 경우, 연구자가 입력한 결정 요인별 제한 범위 내에 해당하는 저널에 대해 필터링한다. 하지만 특정 결정 요인에 대해 연구자 본인에게 적합한 명확한 수치 범위를 알기 어렵다는 점, 결정 요인 별로 중요도가 다르다는 점에서 QoS(quality of service)를 고려하기 위한 추천 서비스 목적[8,9]을 달성하지 못하는 경우가 발생한다.

결정 요인의 중요도를 모두 동일하게 부여하는 일반적

(그림 1) Elsevier에서 제공하는 Journal Finder의 입력 페이지로, 논문 콘텐츠와 추가 수치적 필터링을 통해 논문에 적합한 저널을 추천해주는 서비스.

(Figure 1) Journal recommendation system serviced by Elsevier. Appropriate journals are recommended by analyzing the contents of paper and additional filtering.

인 추천 서비스와 달리, 실제로 고객의 각 결정 요인에 대한 중요도는 상이하고 이를 고려한 추천 시스템에 대한 연구는 다양하게 진행되고 있다 [10,11,12]. 그 중, 요리에 대한 고객의 선호도를 기반으로 요리를 구성하는 재료, 맛, 조리 시간 등의 요인에 대한 고객의 중요도를 계산, 이를 기반으로 요리 조리법을 추천해주는 연구가 존재한다[13]. 이는 고객을 대표할 수 있는 텍스트 등의 구체적인 콘텐츠를 가지고 있지 않아도 선호도만으로 고객에게 걸맞은 추천을 해 줄 수 있다는 점, 추천을 하는 과정에서 결정 요인 별로 상이한 중요도를 고려할 수 있다는 점에서 의의가 있다.

따라서 본 연구에서는 연구자의 과거 논문 게재 이력을 기반으로 저널 결정 요인 별 중요도를 학습하고, 이를 고려한 저널 추천 방법론을 제안한다. 먼저 연구자의 과거 저널 게재 이력을 이용하여 게재한 저널은 선호한다고 가정하여 저널 선호 이력을 형성한다. 저널 선호 이력을 기반으로 연구자 별 각 저널 결정 요인에 대한 중요도를 학습한다. 학습된 중요도 및 저널의 요인 별 값을 이용하여 저널 선호 점수를 계산, 저널 추천을 진행한다.

본 논문의 구성은 다음과 같다. 2장에서는 연구자의 저널 게재 이력 기반 결정 요인 별 중요도를 이용한 저널 추천 방법론을 제시한다. 3장에서는 실제 데이터를 이용한 실험 및 실험 결과에 대해 기술하고, 마지막 4장에서는 결론 및 향후 연구에 대해 다룬다.

2. 제안 방법론

2.1 개요

본 연구는 기존 저널 추천 과정에서 연구자마다 다른 저널 결정 요인 별 중요도를 고려하지 못한다는 점을 해결하기 위해, 저널 게재 이력을 기반으로 결정 요인 별 중

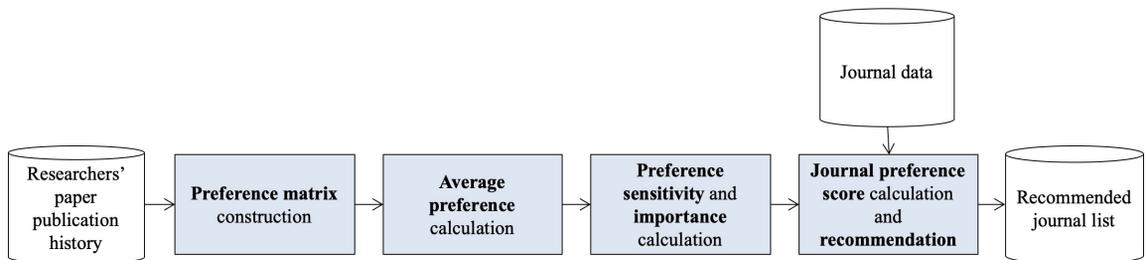
요도를 학습한다. 그림 2는 제안하는 저널 추천 방법론의 전체 개요를 나타낸다. 제안 방법론은 크게 네 단계로 나눌 수 있다.

먼저, 연구자의 저널 게재 이력을 기반으로 연구자의 저널에 대한 선호도 행렬을 형성한다. 다음으로, 이를 이용하여 결정 요인 별 모든 연구자의 평균 선호도를 계산한다. 평균 선호도를 이용하여 각 연구자의 선호 민감도를 계산, 중요도를 학습한다. 마지막 단계에서는 중요도와 저널의 결정 요인별 값을 이용해 선호 점수를 계산하여 최종적으로 적절한 저널을 추천한다.

연구자 별 저널 결정 요인에 대한 중요도를 학습하기 위해 연구자의 논문 게재 이력을 이용한다. 이때, R 은 연구자 집합, J 는 저널 집합을 의미한다. 구체적으로, r_i 는 i 번째 연구자, j_j 는 j 번째 저널을 의미한다.

2.2 선호도 행렬 형성

본 단계에서는 연구자의 연구 논문 게재 이력을 기반으로 저널에 대한 선호도 행렬 P 를 구성하는 과정에 대해 서술한다. 선호도 행렬 P 는 p_{ij} 로 이루어져 있는데 이는 r_i 가 j_j 를 선호하는 정도를 의미하며, 1부터 5의 범위로 이루어져 있다. 선호도 행렬 구성을 위해 과거 논문 게재 이력을 활용한다. 연구자가 과거에 자신의 연구 논문을 특정 저널에 투고하고 게재되었다는 것은 그 저널과 논문이 주제적으로 적합하며, 연구자의 저널의 세부 결정 요인에 대해 만족한다는 것을 의미한다. 따라서 연구자의 논문이 특정 저널에 한 번 게재된 경우 3점, 두 번 이상 게재된 경우 5점을 부여하며, 이는 수식 (1)에 서술되어 있다.



(그림 2) 저널 결정 요인 별 중요도를 고려한 저널 추천 방법론 개요

(Figure 2) Overview of the proposed journal recommendation method considering the importance of each journal decision factor

$$p_{ij} = \begin{cases} 3, & \text{if one paper of } r_i \text{ was published on } j_j, \\ 5, & \text{if two or more paper of } r_i \text{ were published on } j_j. \end{cases} \quad (1)$$

2.3 평균 선호도 계산

전 단계에서 형성한 선호도 행렬 P 를 기반으로 결정 요인별 평균 선호도가 계산되는데, 이를 A_k 로 표현하며 k 번째 요인의 평균 선호도를 나타낸다. 결정 요인 데이터의 형식에 따라 평균 선호도 계산 방법이 나뉜다. 결정 요인의 데이터 형식은 수치 요인과 문자 요인으로 구분하였는데, 수치 요인이라 함은 결정 요인의 값이 연속형 변수인 경우를 의미하고, 문자 요인의 경우 의미를 가지는 단어로 이루어진 경우를 뜻한다. 수치 요인의 경우 특정 수치 요인에 대해 연구자 별로 계산한 결정 요인 값과 선호도 값의 곱의 평균을 계산하여 수치 요인별로 평균 선호도를 얻을 수 있으며, 이는 수식 (2)와 같이 계산할 수 있다.

$$A_k = \frac{\sum_{i=1}^I \sum_{j=1}^J p_{ijk} v_{jk}}{I} \quad (2)$$

이때, p_{ik} 는 r_i 의 k 번째 요인에 대한 선호도를 의미하며 v_{jk} 는 j_j 의 k 번째 요인에 대한 값을 의미, I 는 연구자의 수를 의미한다.

문자 요인의 경우 단어에 대한 평균 선호도를 계산하는데, 모든 저널에 대해서 단어가 특정 저널을 나타내는 값에 포함되어 있으면 그 저널에 대한 연구자의 선호도를 합하여 평균을 구하며, 이를 A_t 로 칭한다. 이때 t 는 단어를 의미한다.

2.4 선호 민감도 및 중요도 계산

선호 민감도는 평균 선호도와 연구자의 선호도 간의 편차율로 정의할 수 있다. 평균 선호도 값 대비 연구자가 특정 요인 혹은 단어에 대해 얼마나 더 혹은 덜 선호하는지 알 수 있다. 수치 요인의 선호 민감도는 수식 (3)과 같이 계산 가능하다.

$$S_{ik} = \frac{|p_{ik} - A_k|}{\min_{l \in K_r, K_j} |l - A_k|}, p_{ik} = \frac{\sum_{j \in J} p_{ij} v_{jk}}{|J|} \quad (3)$$

문자 요인의 경우 단어에 대한 선호 민감도를 기반으로, 특정 요인에 포함된 단어에 대한 값들을 기반으로

선호 민감도를 수식 (4)와 같이 계산할 수 있다.

$$S_{ik} = \frac{\sqrt{\sum_{t \in T_k} (p_{tk} - A_t)^2}}{\sqrt{\sum_t \min_{l \in K_r, K_j} |l - A_k|}}, p = \frac{\sum_{j \in J} p_{ij} v_{jk}}{J} \quad (4)$$

모든 결정 요인에 대한 특정 연구자의 선호 민감도가 계산된 후, 연구자의 각 요인에 대한 중요도를 얻을 수 있으며, 수식 (5)와 같이 계산 가능하다.

$$w_{ik} = \frac{S_{ik}}{\sum_{k \in V_r} S_{ik}} \quad (5)$$

이때, w_{ik} 는 r_i 의 k 번째 요인에 대한 중요도를 의미한다.

2.5 저널 별 선호점수 계산

본 단계에서는 전 단계에서 얻은 요인 별 중요도 및 저널의 요인 별 값을 기반으로 연구자의 모든 저널에 대한 선호 점수를 계산한다. 추가적으로 이를 기반으로 오름차순 정렬하여 저널에 대한 순위를 부여하여 추천이 가능하다. 특정 연구자에 대한 저널 별 선호 점수 F_{ij} 를 계산하는데, 이는 수식 (6)과 같다.

$$F_{ij} = \sum_{k=1}^I \sum_{j=1}^J w_{ik} v_{jk} \quad (6)$$

3. 실험

3.1 데이터 셋

제안 방법론의 성능 평가를 위해 실제 연구자의 논문 게재 이력 및 논문 데이터를 이용해 실험하였다. 총 10명의 연구자로부터 논문 게재 이력을 수집하였으며, 특히 분야 별 평가를 위해 다섯 개의 분야에 걸쳐 연구자를 결정하였다. 데이터가 수집된 연구 분야는 다음과 같다: computer science, business, psychology, medicine, bioscience. 가중치 학습을 위한 결정 요인은 수치 요인 8가지와 문자 요인 3가지를 이용하였으며 이는 표 1에 정리하였다. 결정 요인의 경우, 연구자가 저널을 선택할 때 고려하는 요인에 대한 기존 연구 조사에 따라 선정하였다[14].

(표 1) 실험에 이용한 수치 요인 및 문자 요인의 세부 항목
(Table 1) Summary of numeric factors and simple textual factors which were used for the experiment

데이터 타입	내용
수치 요인	IF, SNIP, review speed, acceptance rate, articles per year, open access, embargo period, publication time and open access fee
문자 요인	Title, abstract, aims & scope

3.2 실험 설계

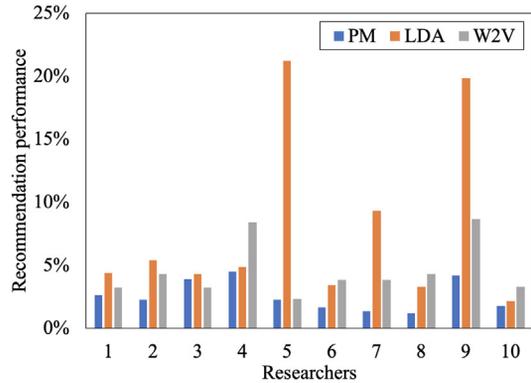
제안 방법론의 성능을 평가하기 위해 두 가지 비교 방법론을 이용하였다. 이는 저명한 텍스트 분석 방법론인 LDA(Latent Dirichlet Allocation), word2vec 방법론이다. LDA 방법론은 텍스트 데이터 기반 토픽 모델링을 통해, 각 문서의 주제간 유사도를 계산하는 방법론이다[15]. 추천 시스템의 성능 평가를 위해 빈번히 사용된다[16,17]. LDA 방법론의 변수로는 주제 모델링 시 생성할 주제의 수, 고려하는 단어의 수가 있는데, 전자를 10부터 100까지 10 단위로, 후자를 10,000부터 50,000까지 10,000 단위로 실험을 진행하였으며 가장 높은 성능을 보인 50, 50,000의 경우를 최종 변수로 설정하였다.

word2vec은 단어 의미 유사성을 고려해 하나의 단어를 벡터 공간에 할당, 이에 대해 관련성을 표현하는 방식을 사용하는 방법론이다[18, 19]. 이용되는 변수로는 결과의 차원 수를 결정하는 값이 있는데, 기존 연구에 따르면 차원 수가 300을 초과하여 커진 경우 정확도의 증가가 미미하나 높은 수준의 계산력을 요구한다고 알려져 있다[20], 따라서 본 논문에서는 차원 수를 300으로 지정하여 실험을 수행하였다.

성능 비교 척도로는 입력 데이터로 사용된 논문이 실제로 투고된 저널이 추천된 순위를 이용하였다. 추천된 순위를 비율로 나타내었으며, 순위가 낮을수록 높은 정도로 추천하는 것을 의미하므로 성능 평가 결과 값이 작을수록 방법론의 성능이 좋음을 알 수 있다.

3.3 실험 결과

제안 방법론의 성능 평가를 위한 실험 결과, 연구자 총 10명에 대해 모든 경우에서 비교 방법론 대비 제안 방법론의 성능이 우수함을 확인하였다. 이는 그림 3을 통해



(그림 3) 제안 방법론 및 비교 방법론을 적용한 실험 결과.
(Figure 3) Recommendation performance obtained by using the proposed and comparative methods.

확인할 수 있는데 ‘PM’은 제안 방법론 기반의 추천을 의미하며, ‘LDA’는 LDA 방법론 기반 주제 유사도 기반의 추천, ‘W2V’은 word2vec 방법론 기반 주제 유사도 추천을 진행한 것을 의미한다. 제안 방법론의 평균 성능은 2.57%, 비교 방법론 중 LDA를 적용한 평균 성능은 7.82%, word2vec 적용 시 4.54%로, 제안 방법론을 적용해 저널 추천을 한 결과가 비교 방법론 대비 높은 순위에 추천되었다.

특히 연구 분야 중 ‘medicine’의 경우 타 분야 대비 최대 3%의 성능 차이를 보이는 것을 확인할 수 있었다. 해당 분야의 경우 평균 IF가 타 분야에 비해 높고, 전문 용어의 비율이 높다는 점을 들어 해당 분야의 특성 파악이 비교적 용이한 정맥 따라 성능이 높은 원인을 파악하였다.

4. 결론

본 연구는 연구자의 논문 게재 이력을 기반으로 저널 결정 요인 별 중요도를 이용한 저널 추천 방법론을 제안한다. 일반적으로 연구자의 저널에 대한 선호 데이터를 얻기 어렵기 때문에, 게재 이력을 이용해 선호도 행렬을 형성하였다. 널리 알려진 IJS의 경우, 토픽 유사도만을 고려한다는 점, 결정 요인을 고려한다 하더라도 결정 요인 별 중요도는 고려하지 않는다는 점에서 한계가 있다. 본 논문에서는 이를 보완하여, 연구자의 게재 이력을 통해 저널 결정 요인 별 중요도를 학습했다는 점에서 의의가 있다. 제안 방법론의 성능을 평가하기 위해 실제 데이터

셋을 이용하여 실험을 진행하였으며, 실험 진행 결과 주제 유사도를 고려하는 비교 방법론에 비해 성능이 우수함을 확인하였다.

향후 연구로는 본 연구에서 다룬 문자 요인을 확장시켜, 단어 단위로 토픽 모델링을 수행한 후 각 토픽(주제)에 대한 선호 민감도를 계산할 수 있다. 제안 방법론과 주제 유사도 기반의 추천 방법론의 양상불을 통해 보다 정교한 모델을 구성할 수 있다. 제시한 프레임워크는 저널 추천뿐 아니라, 텍스트와 결정 요인으로 이루어진 타 정보의 예측에도 확장시킬 수 있을 것이다.

참고문헌(References)

- [1] “Clarivate Analytics - Discover, Protect and Commercialize New Ideas, Faster”, Clarivate Analytics, accessed Apr. 04, 2019, <https://clarivate.com>.
- [2] Ning Kang, Marius A. Doornenbal, Robert J.A. Schijvenaars, “Elsevier Journal Finder: Recommending Journals for your Paper,” Proceedings of the 9th ACM Conference on Recommender Systems, pp.261-264, 2015.
- [3] Camilo Lozoya, Alberto Aguilar-Gonzalez, Antonio Favela-Contreras and Arturo Zamora, “Novus-io: An Internet of Things Platform for Academic Projects”, Transactions on Internet and Information Systems, Vol.12, No.12, pp.5634-5653, 2018.
- [4] Reiswig, Jennifer, “Mendeley,” Journal of the Medical Library Association, Vol.98, No.02, pp.193-194, 2010. <http://dx.doi.org/10.3163/1536-5050.98.2.021>
- [5] Yeonbin Son, Hyeontae An, Yerim Choi, “A Proofreader Matching Method Based on Topic Modeling Using the Importance of Documents”, Journal of Internet Computing and Services, Vol.19, No.04, pp.27-33, 2018. <http://dx.doi.org/10.7472/jksii.2018.19.4.27>
- [6] “Elsevier journal finder”, Elsevier Inc., accessed Apr. 04, 2019, <https://journalfinder.elsevier.com>.
- [7] “Springer Journal Suggester”, Springer Nature, accessed Apr. 04, 2019, <https://journalsuggester.springer.com>.
- [8] Mingchu Li, Xing Jin, Cheng Guo, Jia Liu, Guanghai Cui and Tie Qiu, “RIMNet: Recommendation Incentive Mechanism based on evolutionary game dynamics in peer-to-peer service networks,” Knowledge-Based Systems, Vol.166, pp.156-169, 2019. <https://doi.org/10.1016/j.knosys.2018.12.024>
- [9] Jianxun Liu, Mingdong Tang, Zibin Zheng, Xiaoqing Liu and Saixia Lyu, “Location-aware and personalized collaborative filtering for web service recommendation,” IEEE Transactions on Services Computing, Vol.09, No.05, pp.686-699, 2015. <http://dx.doi.org/10.1109/TSC.2015.2433251>.
- [10] Forbes, Peter, and Mu Zhu, “Content-boosted Matrix Factorization for Recommender Systems: Experiments With Recipe Recommendation.” RecSys, Vol.11, pp.23-27, 2011. <http://dx.doi.org/10.1145/2043932.2043979>
- [11] van Pinxteren, Youri, Gijs Geleijnse, and Paul Kamsteeg, “Deriving a Recipe Similarity Measure for Recommending Healthful Meals”, Proceedings of the 16th International Conference on Intelligent User Interfaces, pp.105-114, 2011. <http://dx.doi.org/10.1145/1943403.1943422>
- [12] Yi-Fan Wang, Yu-Liang Chuang, Mei-Hua Hsu and Huan-Chao Keh, “A Personalized Recommender System for the Cosmetic Business,” Expert Systems with Applications, Vol.26, No.03, pp.427-434, 2004. <https://doi.org/10.1016/j.eswa.2003.10.001>
- [13] Tao, Longquan, Jinli Cao, and Fei Liu, “Dynamic Feature Weighting Based on User Preference Sensitivity for Recommender Systems,” Knowledge-Based Systems, Vol.149, pp.61-75, 2018. <https://doi.org/10.1016/j.knosys.2018.02.019>
- [14] Jan Bröchner and Bo Christer Björk, “Where to Submit? Journal Choice By Construction Management Authors,” Construction Management and Economics, Vol.26, No.07, pp.739-749, 2008. <https://doi.org/10.1080/01446190802017698>
- [15] David M. Blei, Andrew Y. Ng, Michael I. Jordan, “Latent Dirichlet Allocation,” Journal of Machine Learning Research, pp.993-1022, 2003. <https://doi.org/10.1162/jmlr.2003.3.4-5.993>
- [16] Ma, J., Zhang, Y., Wang, Z., &Chen, B., “A New

- Fine-grain SMS Corpus and Its Corresponding Classifier Using Probabilistic Topic Model,” Transactions on Internet and Information Systems, Vol.12, No.02, pp.602-625, 2018.
<http://www.itiis.org/digital-library/manuscript/file/1919/TIIS+Vol+12,+No+2-4.pdf>
- [17] Xinguang Xiang, Fan Liu, Ye Bi, Yanfang Wang and Jinhui Tang, “Local Similarity based Discriminant Analysis for Face Recognition,” Transactions on Internet and Information Systems, Vol.09, No.11, pp.4502-4518, 2015.
<https://doi.org/10.3837/tiis.2015.11.014>
- [18] Tomas Mikolov, Kai Chen, Greg Corrado and Jeffrey Dean, “Efficient Estimation of Word Representations in Vector Space,” arXiv preprint arXiv:1301.3781, 2013.
<https://arxiv.org/abs/1301.3781>
- [19] Joonseo Yun, Hyeontae An, Yerim Choi, “A Machine Learning Based Facility Error Pattern Extraction Framework for Smart Manufacturing”, Journal of Society for e-Business Studies, Vol.23, No.02, pp.97-110, 2018.
<https://doi.org/10.7838/jssebs.2018.23.2.097>
- [20] Xiao, Y., Shi, Q, “Research and Implementation of Hybrid Recommendation Algorithm Based on Collaborative Filtering and word2vec,” Proceedings of the 8th International Symposium on Computational Intelligence and Design, Vol.02, pp.172-175, 2015.
<http://dx.doi.org/10.1109/ISCID.2015.211>

● 저 자 소개 ●



손 연 빈(Yeonbin Son)

2018년 경기대학교 산업경영학과(공학사)
2017년~현재 Desserting Inc. (SUPLE) 대표
2018년~현재 경기대학교 산업경영공학과 석사과정
관심분야 : 머신러닝, 추천시스템
E-mail: yeonbin517@gmail.com



장 태 우(Tai-Woo Chang)

1995년 서울대학교 산업공학과(공학사)
1997년 서울대학교 산업공학과(공학석사)
2004년 서울대학교 산업공학과(공학박사)
2002년~2007년 한국전자통신연구원 연구원/선임연구원
2007년~현재 경기대학교 산업경영공학과 교수
2017년~현재 경기대학교 지능정보융합제조연구센터(GRRC) 센터장
관심분야 : 스마트공장, 지능정보기술, 시스템분석
E-mail: keenbee@kgu.ac.kr



최 예 립(Yerim Choi)

2010년 서울대학교 산업공학과(공학사)
2016년 서울대학교 산업공학과(공학박사)
2016년~2017년 네이버랩스 Data Scientist
2017년~현재 경기대학교 산업경영공학과 조교수
2019년~현재 ai.m Inc. 대표이사
2019년~현재 경기대-KT 빅데이터센터 부센터장
관심분야 : 인공지능/머신러닝, 빅데이터 기반의 인간 모델링
E-mail: yrchoi@kgu.ac.kr