Journal of KIIT. Vol. 23, No. 3, pp. 61-66, Mar. 31, 2025. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2025.23.3.61

이미지 AI 분석을 통한 RAG 기반 실시간 상품 추천 서비스

홍세민*[1], 송주희*[2], 유석종**

RAG-based Real-Time Item Recommendation through Image AI Analysis

Se-Min Hong*[1], Ju-Hee Song*[2], and Seok-Jong Yu**

요 약

현대 사회에서 개인 맞춤형 소비 트렌드가 확산됨에 따라 기업은 소비자의 다양한 요구를 충족하기 위하여 더욱 다양한 선택지를 제공하는 상품들을 생산하고 있다. 그러나 선택지가 많아질수록 소비자는 선택 과정에서 어려움을 겪게 되는 모순이 발생한다. 이러한 문제를 개선하기 위하여 본 연구에서는 이미지 분석과 검색 증강 생성(RAG) 기반 실시간 상품 추천 서비스를 제안한다. 제안 시스템은 GPT-4o 모델을 사용하여 이미지를 인식한 후 RAG를 활용하여 인식된 선택지 중 가장 적합한 선택지 추천을 목표로 한다. 본 시스템에서 선택한 GPT-4o는 기존 광학 문자 인식(OCR) 모델과 비교하여 문맥적 이해도와 수행시간에서 더 우수한 성능을 보였으며, GPT-4o 모델의 상품 라벨 개수에 따른 인식 정확도를 평가하여 최대 라벨 인식 한계 실험 결과를 제시하였다. 이를 통해 소비자는 신속하고 정확하게 추천 상품을 찾을 수 있으며, 향후 AI 기반 라벨 인식률 연구에 기여할 것으로 기대한다.

Abstract

As personalized consumption trends spread in modern society, companies are producing products that provide more diverse choices to meet the diverse needs of consumers. However, as the number of choices increases, consumers experience difficulties in the selection process, which creates a contradiction. Therefore, in this study, we propose a real-time product recommendation service based on image analysis and Retrieval-Augmented Generation(RAG). The proposed system recognizes an image using the GPT-4o model, and then utilizes RAG to suggest the most appropriate choice among the recognized choices. By comparing the existing Optical Character Recognition(OCR) model and the GPT-4o model, we selected GPT-4o, which is superior in contextual understanding and execution time, and presented the maximum label recognition limit by evaluating the recognition accuracy according to the number of labels of the GPT-4o model. This allows consumers to quickly and accurately find the optimal choice and is expected to contribute to improving AI-based label recognition rate in the future based on the results of label recognition experiments.

Keywords

image analysis, GPT-4o model, RAG, text recognition accuracy, recommendation system

* 숙명여자대학교 소프트웨어학부 학사과정
- ORCID[1]: https://orcid.org/0009-0001-0207-3698
- ORCID[2]: https://orcid.org/0009-0008-6407-7353
** 숙명여자대학교 소프트웨어학부 교수(교신저자)
- ORCID: https://orcid.org/0000-0002-1631-4034

· Received: Jan. 06, 2025, Revised: Mar. 06, 2025, Accepted: Mar. 09, 2025
· Corresponding Author: Seok-Jong Yu
Dept. of Computer Science, Sookmyung Womens's University,
Cheongpa-ro 100, 47-gil, Cheongpa-ro, Yongsan-gu, Seoul, Korea
Tel.: +82-2-710-9831, Email: sjyu@sookmyung.ac.kr

Ⅰ. 서 론

현대 사회에서 개인 맞춤형 소비 트렌드가 확산됨에 따라 기업은 소비자의 개별적인 요구를 충족하기 위해 더욱 다양한 상품을 생산하고 있다. 이러한 변화는 소비자에게 폭넓은 선택권을 제공하며 다양한 상품을 구매할 기회를 확대한다. 그러나 선택지가 많아질수록 소비자는 선택 과정에서 어려움을 겪게 되고 이로 인해 만족도가 감소할 수 있다는 연구 결과가 있다[1].

추천 시스템은 데이터를 기반으로 선택지를 제시하여 선택 과정에서 발생하는 어려움을 완화하고 소비자 만족도를 높이는 데 효과적이다. 전자상거래 분야에서 AI를 활용한 추천 시스템이 소비자 맞춤형 서비스를 제공하며 그 효과가 입증되고 있다[2].

구글 렌즈(Google lens)는 이미지를 분석하고 유사 이미지를 제공하여 정보 탐색 방식을 혁신하였으나[3][4], 소비자가 결정을 내리는 데 필요한 다양한 정보를 종합적으로 제공하는 데는 한계가 있다.

본 연구는 이러한 구글 렌즈의 한계를 개선하기 위하여 이미지 분석을 통해 상품 정보를 인식하고 온라인에서 속성 정보를 수집 및 통합하여 소비자가 여러 제품을 비교하고 최적의 선택을 할 수 있도록 지원한다. 이 시스템은 상품 및 도서 추천 등 다양한 분야에 활용될 수 있다.

본 논문은 2장에서 관련 연구를 검토하고 3장에서 시스템의 구현 및 구조를 설명한다. 4장에서는 성능 평가와 분석을 다룬다. 마지막으로 5장에서는 본 연구의 학문적 의의와 한계를 논의하며 연구를 마무리한다.

Ⅱ. 관련 연구

2.1 이미지 분석 기술

이미지 분석 기술은 최근 AI 및 딥러닝의 발전과 함께 다양한 분야에서 널리 활용되고 있다. 이미지에서 정보를 추출하기 위해 YOLO(You Only Look Once)와 광학 문자 인식(OCR, Optical Character Recognition)을 통합하여 ID 배지 이미지에서 정보를 자동으로 추출한 연구가 있다[5]. OCR은 이미지에서 텍스트를 추출하는 기술이다. OCR 기반 기술인 PP-OCR은 텍스트 탐지, 텍스트 인식을 포함한 전체 파이프라인을 지원하며 다국어 지원과 경량화된 구조로 다양한 산업 환경에서 활용되고 있다[6].

OCR은 텍스트 추출에 특화되어 있으나 문맥 이해에는 한계가 있다. 반면, 생성형 AI 모델인 GPT-4o는 이미지에서 텍스트를 추출하고 문맥적 의미 파악, 복잡한 정보 처리에 뛰어난 성능을 보인다[7]. 또한, 라벨 개수에 따른 PP-OCR과 GPT-4o의 라벨 정보 추출 시간을 측정하는 실험을 진행하였고 표 1은 수행시간을 비교한 결과이다.

표 1의 결과에 따르면, OCR은 라벨 수가 증가할수록 처리 시간이 증가하지만 GPT-4o는 일정한 속도로 정보를 처리한다. 본 연구는 이미지를 신속히 분석하여 최적의 선택지를 추천하는 것을 목표로 한다. 따라서 선택지 수와 무관하게 이미지를 빠르게 분석할 수 있는 시스템이 필요하며 이를 위해 GPT-4o를 활용하였다.

표 1. 라벨 개수별 OCR과 GPT-4o 수행시간(초)
Table 1. OCR and GPT-4o performance time by number of labels(second)

Number of labels	OCR execution time	GPT-4o execution time
1	6.7	4.5
2	11.6	3.9
3	15.6	3.9
4	18.3	4.8
5	21.6	5.1
6	26.0	5.5
7	26.7	6.0
8	24.2	7.5

2.2 RAG 기반 응용 기술

RAG(Retrieval-Augmented Generation)는 자연어 처리(NLP, Natural Language Processing) 분야에서 사용되며 정보 검색과 생성을 결합하는 기술이다. RAG를 사용한 대규모 언어 모델(LLM, Large Language Model)은 특정 질문에 답하기 위하여 관련 정보를 검색하고 그 정보를 활용하여 상세하고 정확한 답변을 생성할 수 있다[8].

RAG와 NLP 기술을 활용한 연구에서는 의미 검색, FAISS(Facebook AI Similarity Search) 기반 유사도 매칭, 미세 조정된 GPT-2를 결합한 개인화된 일자리 추천 시스템을 제안한다. 시스템에서 의미 검색과 RAG의 통합은 기존 키워드 매칭 방식보다 직무에 더 적합한 결과를 도출하며 추천 정확도와 사용자 만족도를 향상시켰다[9].

본 연구에서는 타빌리(Tavily) 검색 엔진을 활용하여 실시간 정보를 수집하고 최적의 선택지를 추천하도록 설계하였다. Tavily는 단일 API(Application Programming Interface) 호출당 최대 20개 사이트를 집계하고 독점 AI를 사용하여 쿼리와 가장 관련성이 높은 소스와 콘텐츠를 평가한다[10]. 본 연구는 Tavily를 활용하여 정확도 높은 최신 정보를 제공하고 검색의 효율성을 높이는 데 중점을 두었다. 이는 기존 연구에서 정적 데이터셋을 사용한 것과 달리 실시간 정보 반영이 가능하다는 점에서 차별화된다.

III. 이미지를 통한 RAG 기반 추천 시스템

3.1 시스템 구조

그림 1은 본 연구에서 제안하는 전체 시스템 구조와 처리 과정이다.

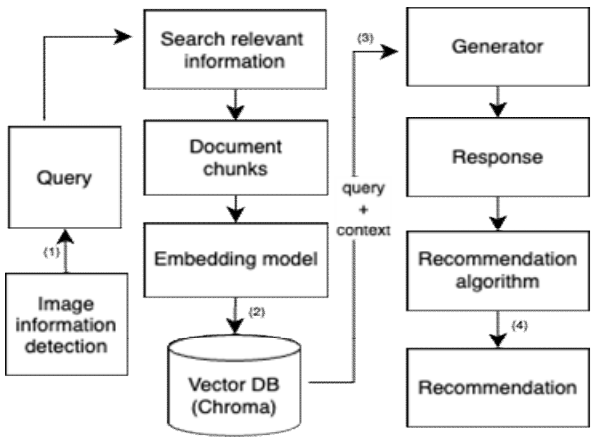

그림 1. 이미지를 통한 RAG 기반 추천 시스템 구조
Fig. 1. RAG-based recommendation system structure using images

(1) 사용자 이미지를 GPT-4o 모델로 분석하여 필요한 정보를 인식한다.
(2) 인식된 정보를 바탕으로 온라인에서 속성 정보를 수집한다.
(3) 저장된 정보 중 관련도가 높은 문서를 기반으로 질문의 답변을 생성한다.
(4) 생성된 답변을 추천 알고리즘에 적용한다.

3.2 RAG 기반 인식 객체의 속성 정보 수집

이미지에서 특정 객체를 인식한 후, 인식된 정보를 바탕으로 관련된 속성 정보를 수집한다. 이 과정은 RAG 기반으로 수행되며 속성 정보를 수집하기 위해 LLM에 최적화된 Tavily 검색 엔진과 GPT-4o를 활용하여 실시간 검색을 진행하였다.

기존 LLM의 생성 기능은 과거 데이터를 학습했기 때문에 최신 정보가 부족하여 발생하는 환각 현상이 가장 큰 문제이다. 이를 해결하기 위해 도입된 RAG는 사용자가 제공한 벡터 데이터베이스를 기반으로 응답을 생성하기 때문에 존재하지 않는 정보를 생성하는 환각 현상을 줄이는데 효과적이다[11].

본 연구에서는 Tavily로 질의에 따른 최신 온라인 속성 정보를 수집하고, 이를 기반으로 응답을 생성하도록 하였다. Tavily로 수집된 문서는 텍스트를 추출한 후, LangChain을 활용해 Document 객체로 변환하여 시스템에 저장한다. LangChain은 LLM을 활용한 애플리케이션 개발에 특화된 오픈소스 프레임워크이다[12]. 변환된 Document는 OpenAI의 임베딩(Embedding) 모델로 벡터화하여 Chroma 벡터 데이터베이스에 저장한 후, 검색기(Retriever)를 활용하여 질의와 관련성이 높은 데이터를 검색한다. 검색된 문서는 질의 간의 관련성과 다양성을 조절하는 MMR(Multi-Modal Retrieval)을 기반으로 선별한 후, GPT(Generative Pre-trained Transformer) 모델의 질문에 대한 응답 생성 과정에서 활용된다. 표 2는 MMR과 GPT-4o를 통해 생성한 속성 정보이다.

표 2. MMR에 의해 생성된 속성 정보
Table 2. Attribute information created through MMR

Product name	Number of reviews	Rating	Price
페브리즈 370ML(상쾌한향)	30307	5.0	2090
페브리즈 다우니 실내건조 370ML	4749	4.8	3540
페브리즈 370ML(은은한향)	1634	4.5	4400

3.3 랭킹 추천 알고리즘

본 시스템의 랭킹 추천 알고리즘은 사용자가 인식한 상품들 간에서만 비교를 수행하고, 인식된 상품들의 속성 정보를 검색하여 가장 적합한 선택지를 제공한다. 이는 지식 기반 추천 시스템과 콘텐츠 기반 필터링(Content-based filtering)의 장점을 결합한 접근법으로, 인식된 상품의 평점, 리뷰 수, 가격과 같은 주요 상품 속성을 기반으로 비교를 진행한다. 속성 정보의 비교는 순위 알고리즘을 활용한다. 평점과 리뷰수는 사용자 만족도와 신뢰도를 반영하는 지표로, 값이 높거나 많을수록 높은 순위를 부여한다. 가격의 경우 온라인 최저가와 현장가의 차이가 작을수록 현장에서 바로 구매할 경제적 합리성이 높다고 평가되며, 더 높은 순위를 얻게 된다. 이와 같은 방식은 상품 속성별로 독립적인 평가를 진행한 뒤, 통합적으로 가장 높은 순위를 가진 선택지를 추천함으로써 사용자가 신뢰할 수 있는 합리적인 선택지를 제안한다.

식 (1)은 식 (2)를 바탕으로 모든 상품 i의 최종 순위를 계산한 후, 가장 높은 순위를 가진 상품을 도출한다.

$$f(x) = \arg\ \max_i(S_{final}(i)) \tag{1}$$

$$S_{final}(i) = \sum_{k=1}^{n}(100 - 10 \times (Rank_k(i) - 1)) \tag{2}$$

여기서 $S_{final}(i)$는 상품 i의 최종 순위, $Rank_k(i)$는 상품 i의 속성 k에 대한 순위, n은 평가 항목의 총 개수이다.

인식된 상품들과 유사 상품을 추천하기 위하여 TF-IDF(Term Frequency-Inverse Document Frequency) 기반 코사인 유사도 이외에 OpenAI Embedding을 활용하여 상품 간 의미적 유사성을 분석하였다.

Ⅳ. 시스템 구현 및 성능 평가

4.1 데이터 수집 및 실험 환경

이미지에서 라벨 개수에 따른 문자 인식의 정확도를 평가하기 위해 이마트를 기준으로 가격표 사진 데이터를 직접 수집하였다. 실험 데이터는 1개에서 8개까지 서로 다른 개수의 가격표 라벨을 포함한 사진으로 구성되었으며, 개수별로 문자 인식의 정확도와 상품명 및 가격정보의 매칭 정확도를 분석하였다. 각 개수별로 50장의 사진을 수집하여 총 400장의 사진을 실험 데이터로 활용하였다. 수집된 모든 사진은 가격표의 모서리를 기준으로 크롭하여 데이터셋을 구축하였다. 이후 정답 데이터셋을 작성하고 GPT-4o를 이용한 문자 인식 결과와 비교하여 정확도를 평가하였다. 표 3은 성능 평가에 사용된 데이터셋 이미지를 나타내며 주요 라이브러리는 Tensorflow 2.17.0, OpenCV 4.10.0 버전을 사용하였다.

표 3. 라벨 개수별 데이터셋 이미지
Table 3. Dataset images by number of labels

labels	Image
1	
4	
8	

4.2 성능 평가방법

문자 인식 성능 평가는 WRA(Word Recognition Accuracy), 1-NED(Normalized Edit Distance), 항목 수준 정확도(Item-level accuracy)를 지표로 삼아 수행되었다. WRA는 정답 글자와 예측된 글자를 비교하여 일치 여부를 판단하는 평가방법이다. 하지만 단어 단위로 평가할 경우, 단어 중 맞은 글자 수를 반영하지 못하는 단점이 있다. 이를 보완하기 위해 1-NED 기법이 사용된다. 1-NED 방법은 정답 단어와 예측 단어를 비교하여 문자열 편집 거리(Edit Distance)를 정규화한 값으로, 값이 1에 가까울수록 인식이 정확함을 의미한다[13].

하지만 WRA, 1-NED 모두 상품과 가격 간 매칭 정확도 확인에 직관성이 떨어지므로 본 연구에서는 항목 수준 정확도를 추가로 평가하였다. 항목 수준 정확도는 상품명과 가격정보를 하나의 항목으로 간주한다. 이를 통해 각각의 항목이 정확히 매칭되었는지 확인하여 문자 인식뿐만 아니라 상품명과 가격 간의 연관성을 종합적으로 평가한다.

4.3 성능 평가 결과

그림 2와 같이 문자 인식 정확도는 라벨의 개수가 증가할수록 감소하는 경향을 보였다. 라벨 개수 1개에서 4개까지는 비교적 안정적인 성능을 유지하였으나, 라벨 수가 5개로 증가하면서 WRA는 69.3%, 1-NED는 0.87, 항목 수준 정확도는 38.8%로 급격히 감소하였다. 라벨 수가 8개로 늘어난 경우, WRA는 39.4%, 1-NED는 0.66, 상품명과 가격 매칭 정확도는 11.7%로 성능이 크게 저하되었다.

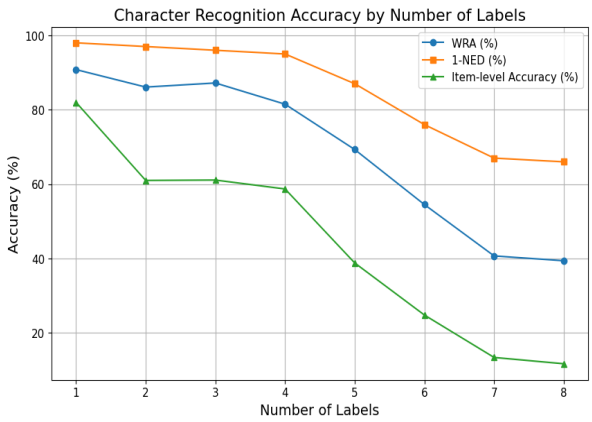

그림 2. 성능 평가 결과
Fig. 2. Experimental results

4.4 추천 결과

그림 3은 Streamlit을 사용하여 개발한 상품 추천 데모 페이지 화면이다. Streamlit은 데이터 시각화에 사용되는 Python 기반의 오픈소스 웹 애플리케이션 프레임워크이다. 또한, Streamlit Cloud를 통해 데모 페이지를 배포하여 웹과 모바일 기기를 통해 URL로 손쉽게 접근할 수 있도록 접근성을 높였다.

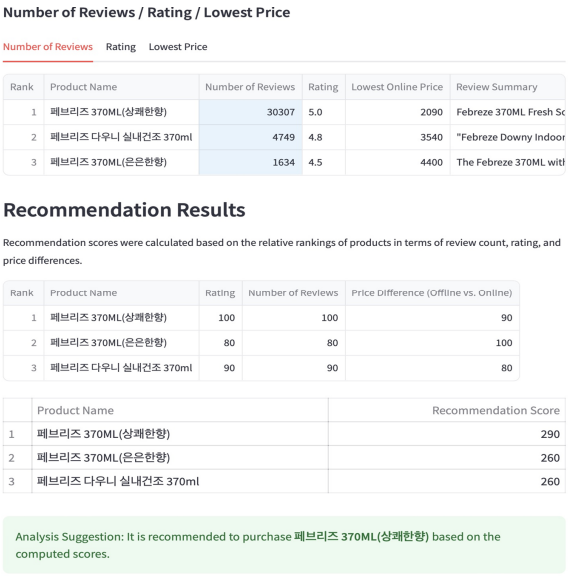

그림 3. 상품 추천 결과
Fig. 3. Product recommendation results

V. 결론 및 향후 과제

본 연구는 기존 텍스트 검색의 단점을 보완하는 이미지 기반 AI 추천 시스템이라는 점에서 의의를 갖는다. 또한, 제안 시스템은 특정 상품군에 국한되지 않고 다양한 환경에서 활용할 수 있는 확장성을 갖추고 있다. 그러나 RAG와 LangChain을 활용한 웹 검색 과정에서 특정 검색어에 대해 결과를 반환하지 못하는 경우가 있었으며, 생성형 AI를 활용한 라벨 인식 과정에서는 라벨의 개수가 많아질수록 인식 성능이 저하되는 문제가 관찰되었다. 또한, 가격표의 디자인이 표준화되지 않아 데이터셋 구축에 제약이 있었다. 향후 연구에서는 다양한 디자인을 포함하는 표준화된 데이터셋을 구축하여 일반적인 환경에서도 높은 성능을 유지하도록 개선할 필요가 있으며, 인식 성능 저하 문제는 AI 모델의 기술적 발전을 통해 극복할 수 있을 것으로 기대한다.

References

[1] S. S. Iyengar and M. R. Lepper, "When Choice is Demotivating: Can One Desire Too Much of a Good Thing?", Journal of Personality and Social Psychology, Vol. 79, No. 6, pp. 995-1006, Dec. 2000. https://doi.org/10.1037//0022-3514.79.6.995.

[2] S. Sinha and M. Rakhra, "AI Driven E-Commerce Product Recommendation", Proc. 2023 6th Int. Conf. on Contemporary Computing and Informatics (IC3I), Gautam Buddha Nagar, India, pp. 13-19, Sep. 2023. https://doi.org/10.1109/IC3I59117.2023.10397621.

[3] How Lens Works, https://lens.google/intl/ko/howlensworks/. [accessed: Feb. 05, 2025]

[4] D. H. Salim, J. J. Susanto, S. R. Manalu, and H. A. Shiddiqi, "Through the Lens: Unveiling the Power and Promise of Google's Visual Search Technology", Proc. 2024 Int. Conf. on Information Management and Technology (ICIMTech), Bali, Indonesia, pp. 207-211, Aug. 2024. https://doi.org/10.1109/icimtech63123.2024.10780912.

[5] W. Cavalcante, I. G. Torné, L. Camelo, R. Fernandes, A. Printes, and H. Bragança, "An ID Badge Information Extractor Based on Object Detection and Optical Character Recognition", IEEE Access, Vol. 12, pp. 152559-152567, Oct. 2024. https://doi.org/10.1109/ACCESS.2024.3471449.

[6] Y. Du, et al., "PP-OCR: A Practical Ultra Lightweight OCR System", arXiv preprint arXiv:2009.09941, Sep. 2020. https://doi.org/10.48550/arXiv.2009.09941.

[7] J. Achiam, et al., "GPT-4 Technical Report", arXiv preprint arXiv:2303.08774, Mar. 2023. https://doi.org/10.48550/arXiv.2303.08774.

[8] J. Y. Seo, "Developing AI Services Based on LLM with LangChain", Gilbut, pp. 47-60, Feb. 2024.

[9] S. Deshmukh and A. Bajaj, "CareerBoost: A Hybrid RAG-NLP Job Recommendation Framework", Proc. 2024 8th Int. Conf. on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), Kirtipur, Nepal, pp. 853-858, Oct. 2024. https://doi.org/10.1109/I-SMAC61858.2024.10714727.

[10] Tavily introduction, https://docs.tavily.com/guides/introduction. [accessed: Feb. 05, 2025]

[11] N. H. Kim and J. Y. Kang, "Evaluation of Learning Outcomes and Efficiency of Generative AI Using Metadata-Based RAG", The Journal of Society for e-Business Studies, Vol. 29, No. 4, pp. 11-30, Nov. 2024. https://doi.org/10.7838/jsebs.2024.29.4.011.

[12] LangChain introduction, https://www.samsungsds.com/kr/insights/what-is-langchain.html. [accessed: Feb. 05, 2025]

[13] S. H. Sung, K. B. Lee, and S. H. Park, "Research on Korea Text Recognition in Images Using Deep Learning", Journal of the Korea Convergence Society, Vol. 11, No. 6, pp. 1-6, Jun. 2020. https://doi.org/10.15207/JKCS.2020.11.6.001.

저자소개

홍 세 민 (Se-Min Hong)

2022년 2월 ~ 현재 : 숙명여자대학교 소프트웨어학부 학사과정

관심분야 : 생성모델, 프로그래밍

송 주 희 (Ju-Hee Song)

2022년 2월 ~ 현재 : 숙명여자대학교 소프트웨어학부 학사과정

관심분야 : 생성모델, 자연어처리

유 석 종 (Seok-Jong Yu)

1994년 2월 : 연세대학교 전산과학과(이학사)

1996년 2월 : 연세대학교 컴퓨터과학과(이학석사)

2001년 2월 : 연세대학교 컴퓨터과학과(공학박사)

2005년 ~ 현재 : 숙명여자대학교 소프트웨어학부 교수

관심분야 : 데이터마이닝, 추천시스템, 정보시각화