

RAG 과 Content-based Recommendation Systems 을 이용한 메신저 통합 관리 플랫폼

오진우¹, 박서연²

¹ 한양대학교 ERICA 컴퓨터학부 학부생

한양대학교 ERICA 컴퓨터학부 교수

oh7895@hanyang.ac.kr, seoyeonpark@hanyang.ac.kr

An Integrated Messenger Management Platform Utilizing RAG and Content-Based Recommendation Systems

Jin-Woo Oh¹, Seo-Yeon Park²

¹Dept. of Computer Science, Hanyang University ERICA

²Dept. of Computer Science, Hanyang University ERICA

요약

본 연구에서는 소통의 본질을 이루기 위해, 다양한 메신저 플랫폼들을 하나로 통합하고 수신한 메시지들의 알림 여부를 딥러닝과 콘텐츠 기반 추천 시스템을 이용하여 판단한 후 결과에 따라 요약하여 제공하는 “메신저 통합 관리 플랫폼”을 제안한다. 이 과정에서 RAG, Content-based Recommendation System 과 같은 기술들의 이용이 사용자가 앞으로 중요하다고 여길 만한 메시지를 판별하는 과정에서 유의미한 결과를 가져오도록 설계했는지, 따라서 향후 사용자 친화적 중요 메시지 분류에 중요한 역할을 어떤 식으로 할 것인지 설득하는 과정을 소개한다.

1. 서론

스마트폰과 이에 따른 메신저 및 SNS 플랫폼들의 발전으로 인간과 인간 간의 소통의 방식이 매우 다양해졌다. 그로 인해, 각 메신저 플랫폼마다 차별화된 소통 형식으로 각 상황에 맞는 정보를 더욱 효과적으로 볼 수 있는 등 장점들이 있겠지만, 한편으로는 너무 다양하기 때문에 사용자 입장에선 오히려 헷갈리는 상황이 발생하게 된다. 또한, 특정 세대나 성향을 기준으로 메신저 플랫폼이 분리되는 현상도 생겨 메신저의 다양화가 되려 통신의 장애를 일으킬 수 있다 [1].

이에 본 연구에서는 방대한 종류의 메신저들을 하 나의 서비스에서 관리하여 메시지들을 한 눈에 볼 수 있으면서 중요한 정보만 요약해서 제공하는 “메신저 통합 관리 플랫폼”을 제안한다. 이 과정에서 LLM 과 Content-based Recommendation Systems 를 이용하여 사용자가 메신저 플랫폼에 구애받지 않고 필요한 정보만 빠르고 간결하게 습득하는 것, 진정한 의미의 소통을 달성하는 것이 이번 연구의 목표이다.

2. 주요 기능 제안

2-1. 메신저 계정 통합 검색

아래의 [그림 1]에서 확인할 수 있듯이 OAuth2.0 프로토콜을 이용하여 선택한 메신저들의 Access Token 을 받아와 계정을 가져올 뿐만 아니라 API 에도 접근

할 수 있도록 한다. OAuth 를 지원하지 않는 메신저들은 추후에 기술을 개발하여 통합할 예정이다.

2-2. 중요 메시지 분류

사용자가 설정한 키워드와 주제를 기준으로 다양한 메신저에서 온 메시지들의 문맥을 이해하여 사용자에게 추천할 만한 메시지를 선별해내는 기능이다. 이 과정에서 RAG 기술을 통해 외부 지식 및 과거 데이터를 검색하여 특정 도메인 지식을 깊게 이해하고, GPT-4o-mini 로 문맥을 파악하여 메시지의 키워드를 추출한다. 이렇게 추출한 키워드를 벡터화하고 유튜브 알고리즘에서 쓰이는 Content-based Recommendation Systems 을 적용하여 사용자 추천 여부를 결정한다. 이를 통해 어느 메신저에서 온 메시지든 본인이 설정한 중요 내용을 기준으로 메시지를 분류함으로써 필요한 메시지에 쉽게 접근할 수 있게 된다.

2-3. 메시지 핵심 내용 요약

위에서 중요하다고 분류한 메시지의 내용을 다시 한 번 LLM Prompting 을 통해 사용자가 설정한 키워드/주제의 내용들만 선별·요약해서 알림을 해주는 기능이다. 이로써 사용자는 전체 메시지를 일일이 확인하지 않아도 자신에게 필요한 핵심 정보만 빠르게 파악할 수 있다. 특히, 다양한 메신저를 모은다는 서비스 특성상 중요한 메시지를 분류하여 사용자에게 전달하더라도 혼란을 초래할 수 있는데 불필요한 정보를 걸러내고, 사용자의 관심사에 맞는 내용만을 추출해

전달함으로써 중요 내용을 효과적으로 전달한다.

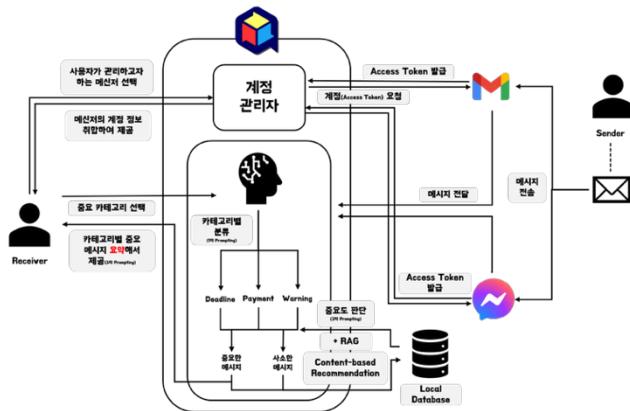


그림 1 : CoChat 시스템 설계도

3. 데이터 처리 흐름

3-1. RAG(Retrieval Augmented Generation)

RAG는 외부 지식 검색과 생성형 AI를 결합해 정확한 응답을 생성하는 기술이다. [2,3,4] RAG 기술의 목적은 LLM의 환각(Hallucination) 감소[5], 도메인 특화 지식 활용[6]에 있다.



그림 2 : RAG 이용 방식

[그림 2]와 같이 새로 들어오는 메시지의 내용에 LLM이 이해하기 어려운 은어인 “플젝”과 같은 단어를 RAG를 통해 의미를 검색/추론하여 메시지의 키워드를 더 정확하게 추출해낸다.

3-2. Content-based Recommendation System

콘텐츠 기반 추천 시스템은 아이템의 특징과 사용자의 과거 행동/선호도를 기반으로 유사한 아이템을 추천하는 방식이다[7]. 이는 사용자가 선호한 아이템의 특징을 분석해 유사한 새 아이템을 매칭한다.

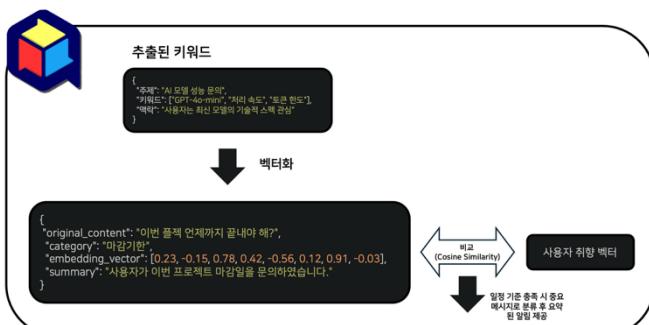


그림 3 : Content-based Recommendation System 이용 방식

[그림 3]과 같이 RAG를 통해 추출한 키워드를 특정 벡터화 알고리즘을 적용하여 벡터화를 진행한 뒤, 사용자의 취향을 반영하고 있는 벡터와 Cosine Similarity로 비교하여 유사도를 구한다. 만약, 유사도

가 일정 수치 이상을 만족하면, 사용자에게 요약해서 알릴 “중요메시지”로 분류되는 것이다.

4. 결론

본 연구는 메신저 플랫폼의 과정화로 인한 소통 과부하 문제를 해결하기 위해 RAG과 Content-based Recommendation Systems을 결합하여 사용자에게 소통의 본질인 “정보 전달”的 확실성을 높이는 과정을 탐구했다. 특히, RAG을 통해 외부 지식과 내부 대화 기록을 실시간 검색함으로써 은어나 고유 명사와 같은 단어에 대한 이해도를 높이고 사용자 취향 벡터와의 코사인 유사도 분석을 통한 중요도 평가로 맞춤형 알림 정확도를 향상시킬 수 있다. 또한, 근본적으로 다양한 메신저 플랫폼을 하나의 앱으로 통합하여 더욱 세대나 성향과 같은 개인적인 특성에 따라 소통이 단절되는 현상을 줄인다. 결론적으로, 본 시스템은 메신저 관리의 효율성과 개인화를 동시에 실현하며, 학술적·실용적 가치를 지닌다.

참고문헌

- [1] Afzal, A., Hussain, M., Saleem, S., Shahzad, M. K., Ho, A. T., & Jung, K. H. (2021). Encrypted network traffic analysis of secure instant messaging application: A case study of signal messenger app. *Applied Sciences*, 11(17), 7789.
- [2] Ram, O., Levine, Y., Dalmedigos, I., Muhlgray, D., Shashua, A., Leyton-Brown, K., & Shoham, Y. (2023). In-Context Retrieval-Augmented Language Models. *Transactions of the Association for Computational Linguistics*, 11, 1316-1331.
- [3] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, 33, 9459-9474.
- [4] Rau, D., Déjean, H., Chirkova, N., Formal, T., Wang, S., Clinchant, S., & Nikoulina, V. (2024, November). BERGEN: A Benchmarking Library for Retrieval-Augmented Generation. In *Findings of the Association for Computational Linguistics: EMNLP 2024* (pp. 7640-7663).
- [5] Niu, C., Wu, Y., Zhu, J., Xu, S., Shum, K., Zhong, R., ... & Zhang, T. (2024, August). RAGTruth: A Hallucination Corpus for Developing Trustworthy Retrieval-Augmented Language Models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 10862-10878).
- [6] Mao, K., Liu, Z., Qian, H., Mo, F., Deng, C., & Dou, Z. (2024, November). RAG-Studio: Towards In-Domain Adaptation of Retrieval Augmented Generation Through Self-Alignment. In *Findings of the Association for Computational Linguistics: EMNLP 2024* (pp. 725-735).
- [7] Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In *The adaptive web: methods and strategies of web personalization* (pp. 325-341). Berlin, Heidelberg: Springer Berlin Heidelberg.