

PBNR : Prompt-based News Recommender System

<https://arxiv.org/pdf/2304.07862>

0. Introduction

- 뉴스 추천은 실시간 관심사 반영이 핵심인 도메인
- 기존 추천 시스템은 학습 주기와 모델 구조에 묶여 변화 대응이 느림
- 임베딩 기반 접근은 새로운 이슈·뉴스에 취약
- 대규모 언어 모델은 자연어 이해와 추론을 통해 추천 패러다임을 확장
- 본 논문은 프롬프트 기반 뉴스 추천 시스템 PBNR을 제안

1. Overview

- 뉴스 추천 문제를 자연어 생성 및 추론 문제로 재정의
- 사용자 이력과 뉴스 정보를 하나의 프롬프트로 통합
- LLM이 추천 대상 선택 또는 순위 생성을 직접 수행
- 별도의 추천 모델 학습 없이 추천 가능
- 빠른 적응성과 설명 가능성을 주요 목표로 설정

2. Challenges

- 뉴스 소비 패턴은 짧은 시간 내 급격히 변화
- 기존 시스템은 재학습 비용과 지연 문제가 큼
- LLM 기반 접근은 추론 비용이 높음
- 프롬프트 설계 품질에 따라 성능 편차 발생
- 추천 결과의 일관성과 재현성 확보가 어려움

3. Method

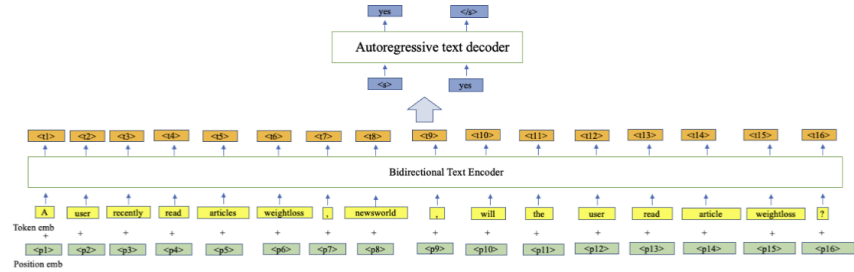


Figure 1: PBNR utilizes an encoder-decoder framework, where a user's historical behavior is converted into a text inquiry and each news article is described textually, and then PBNR generates the answer to indicate a user's preference to a candidate article through an auto-regressive decoder.

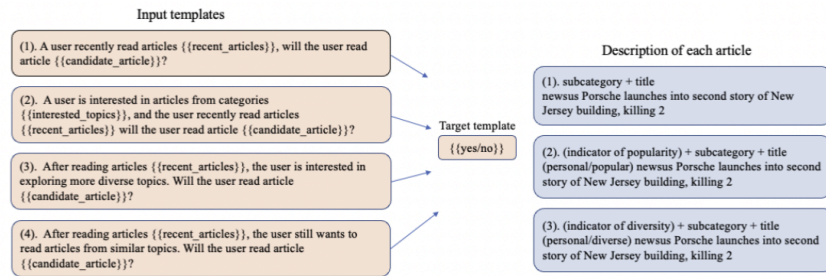


Figure 2: The personalized prompts are created by designing input-target templates, wherein the relevant fields in the prompts are replaced with corresponding information from the raw data. The description of articles may differ based on specific recommendation tasks. In this study, the model denoted by PBNR (i - j) employs input template (i) and article description (j).

- 추천 과정을 질문 응답 형태의 프롬프트로 구성
- 프롬프트는 세 부분으로 구성
 - 사용자 뉴스 소비 이력 요약
 - 후보 뉴스의 콘텐츠 정보
 - 추천 기준 및 출력 형식 지침
- LLM은 프롬프트를 기반으로 추천 뉴스 선택 또는 랭킹 생성
- in-context learning 기반 추론 방식 사용

4. Experiments

- 뉴스 추천 벤치마크 데이터셋 활용
- 기존 뉴스 추천 모델 및 임베딩 기반 방식과 비교
- 추천 성능 평가를 위한 정량 지표 사용

- 프롬프트 구성 변화에 따른 성능 비교 실험 수행

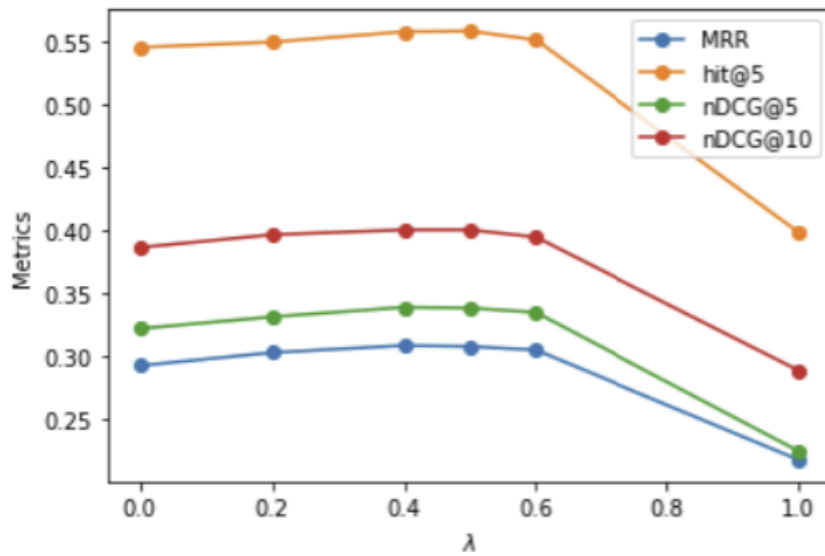
5. Results

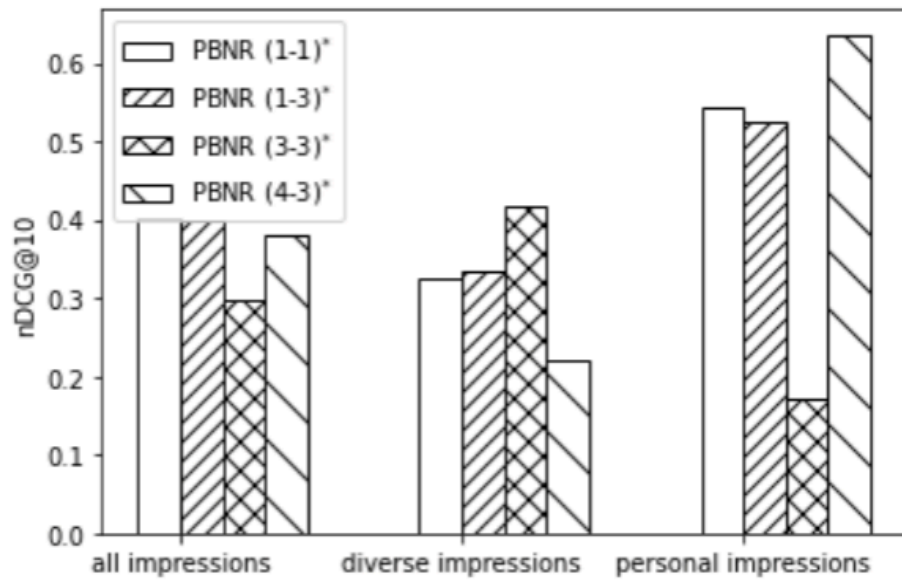
Table 1: Statistics of MIND used for model evaluations.

#users	#news	#impressions	avg. history length	avg. click rate (%)	avg. title length	#category
141,935	71,671	297,715	23.56	0.10	10.77	18

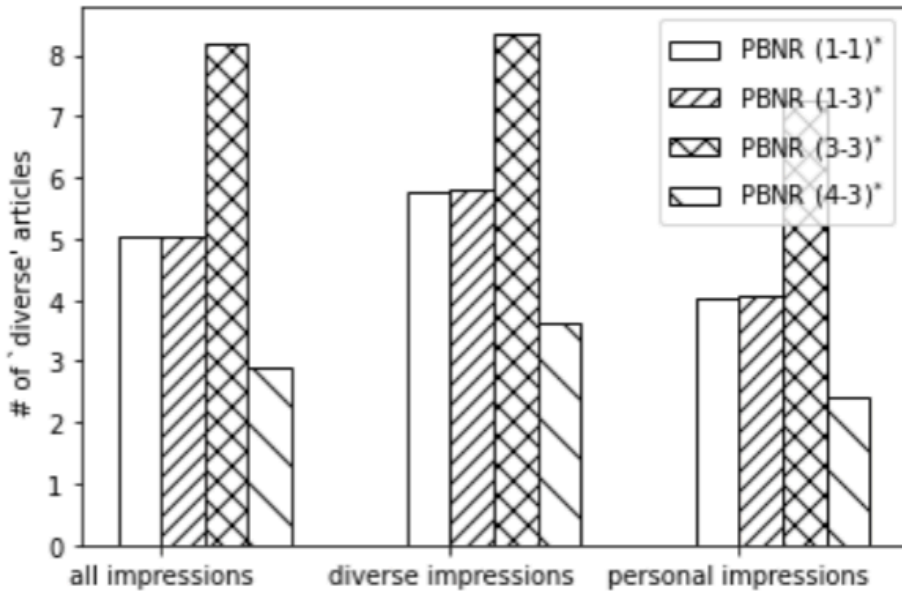
Performances on sequential news RS					Performances on pure cold-start users			
Methods	MRR	HR@5	NDCG@5	NDCG@10	MRR	HR@5	NDCG@5	NDCG@10
MostPop	0.2699	0.4899	0.2906	0.3510	0.2699	0.4899	0.2906	0.3510
RecentPop	0.2704	0.4939	0.2924	0.3519	0.2704	0.4939	0.2924	0.3519
LSTUR	0.2522	0.4715	0.2712	0.3352	0.2143	0.4027	0.2258	0.2842
TANR	0.2918	0.5519	0.3241	0.3876	0.2180	0.3905	0.2232	0.2855
NRMS	0.2847	0.5253	0.3101	0.3763	0.2501	0.4600	0.2669	0.3289
NAML	0.2943	0.5426	0.3235	0.3870	0.2180	0.3905	0.2232	0.2855
PBNR (1-1)	0.2924	0.5450	0.3218	0.3862	0.2308	0.4458	0.2472	0.3107
PBNR (1-1)*	0.3084	0.5574	0.3387	0.4012	0.2469	0.4888	0.2721	0.3322
PBNR (2-1)	0.3048	0.5521	0.3341	0.3971	0.2602	0.4839	0.2790	0.3464

Methods	HR@5	NDCG@5	NDCG@10	Gini@5	Gini@10	Topic@5	Topic@10	New@5	New@10	tradeoff
MostPop	0.4899	0.2906	0.3510	0.6716	0.7675	3.6178	5.8699	2.7593	5.4906	0.4437
RecentPop	0.4939	0.2925	0.3519	0.6695	0.7664	3.5924	5.8734	2.7331	5.4630	0.4447
TANR	0.5519	0.3241	0.3876	0.5934	0.7073	3.2685	5.1822	2.3216	4.6841	0.4601
NAML	0.5426	0.3235	0.3869	0.5989	0.7109	3.2958	5.2412	2.2952	4.6713	0.4605
PBNR (1-1)*	0.5574	0.3387	0.4012	0.5678	0.6972	3.1467	5.1722	1.7172	3.9584	0.4668
PBNR (1-2)*	0.5714	0.3470	0.4076	0.5384	0.6794	2.9884	4.9450	1.6103	3.7911	0.4658
PBNR (1-3)*	0.5688	0.3454	0.4068	0.5464	0.6828	3.0369	4.9535	1.6092	3.7424	0.4666





a Model performances on sequential news recommendation task.



b The number of articles labelled as 'diverse' within the top-10 recommendations.

- PBNR은 기존 추천 모델과 유사하거나 경쟁력 있는 성능을 달성
- 사용자 이력이 부족한 상황에서 상대적 성능 유지
- 신규 뉴스 유입 시 성능 저하가 적음
- 프롬프트 설계에 따라 성능 변동 폭이 큼

- 추론 비용과 응답 시간 증가라는 명확한 트레이드오프 존재

6. Insight

- PBNR의 핵심 기여는 추천 문제의 추상화 수준을 한 단계 끌어올린 점
- 추천 로직을 모델 파라미터가 아닌 자연어 지침으로 외부화
- 뉴스 추천처럼 변화가 잦은 도메인에서 재학습 없는 적응이 가능
- 추천 결과에 대한 설명을 자연어로 함께 제공할 수 있어 해석 가능성 증가
- 성능 자체보다 시스템 유연성과 유지 비용 구조 변화가 더 큰 의미
- 다만 추론 비용과 응답 지연은 대규모 서비스 적용의 주요 병목