# PPT: A Minor Language News Recommendation Model via Cross-Lingual Preference Pattern Transfer

**Yiyang Zhang***, **Nan Chen***
University of Science and Technology of China, Hefei, China
{yiyangzhang, chen_nan}@mail.ustc.edu.cn

## Abstract

Rich user-item interactions are essential for building reliable recommender systems, as they reflect user preference patterns. However, minor language news recommendation platforms suffer from limited interactions due to a small user base. A natural solution is to apply well-established English recommender systems to minor language news recommendation, but the linguistic gap can lead to inaccurate modeling of minor language news content. Therefore, enabling few-shot minor language news recommender systems to capture both content information and preference patterns remains a challenge. Based on the observation that preference patterns are similar across languages, we propose a minor language news recommendation model by cross-lingual preference pattern transfer, named PPT. Our model adopts the widely used two-tower architecture and employs the large language model as the backbone of the news encoder. Through cross-lingual alignment, the strong English capability of the news encoder is extended to minor languages, thus enhancing news content representations. Additionally, through cross-lingual news augmentation, PPT simulates interactions of minor language news in the English domain, which facilitates the transfer of preference patterns from the many-shot English domain to the few-shot minor language domain. Extensive experiments on two real-world datasets across 15 minor languages demonstrate the superiority and generalization of our proposed PPT in addressing minor language news recommendation.

## 1 Introduction

In today's fast-paced world, online news platforms in various languages play a crucial role in keeping individuals informed. With the daily surge of news articles, it is important to develop personalized news recommender systems to help users navigate the overwhelming flow of information.

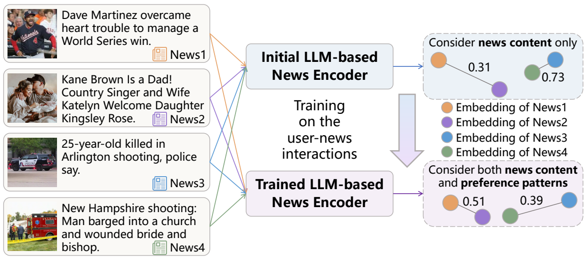



Figure 1: The process of learning preference patterns. The number denotes the cosine similarity. Note that the shorter the distance, the higher the similarity.

A widely adopted paradigm for news recommendation is to separately represent users and news as embeddings and then predict user interest in candidate news (Wu et al., 2023). User embeddings are derived from clicked news embeddings, while news embeddings incorporate both content information and preference patterns. Content information can be directly extracted by the news encoder, whereas preference patterns need to be learned by training on user-news interactions. Figure 1 illustrates the process of learning preference patterns. After training, the similarity between News1 and News2 increases while the similarity between News3 and News4 decreases. News1 and News2 differ significantly in terms of content (a baseball coach overcomes illness to win vs. a country singer welcomes daughter), which causes low initial similarity. However, both News1 and News2 convey positive emotions, and baseball and country music share a common cultural foundation (Cusic, 2003; Vignola, 2005). As a result, they are more likely to appeal to the same user. This preference pattern is learned through training and leads to an increase in similarity. As for News3 and News4, although they are both crime news articles about shooting incidents, readers tend to prefer local or regional news (Schrøder, 2019), a pattern that is also learned by training. Consequently, their similarity decreases since News3 occurred in Arlington,

*Both authors are corresponding authors.



*Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10728–10745
July 27 - August 1, 2025 

Texas, and News4 in New Hampshire.

Therefore, the accuracy of recommender systems relies heavily on interactions, as evidenced by the analysis in Appendix B.1. English news platforms with a large global user base have rich interactions (many-shot), whereas minor language platforms with limited users face interaction scarcity (few-shot). For example, the average number of historical interactions per user on the Norwegian news platform Adressa (Gulla et al., 2017) is only 8.71, far less than the average of 25.27 on English Microsoft News (Wu et al., 2020). In such cases, minor language news recommendation performance is primarily constrained by insufficient preference patterns rather than the news encoder's ability to capture content. An intuitive method is to train the model in the English domain and directly apply it to the minor language domain (Guo et al., 2023), but this method performs poorly due to the linguistic gap. Translating minor language news into English is a natural attempt. However, our preliminary experiments show that while translation improves performance to some extent, the semantic shifts caused by translation distort content representation and hinder preference pattern transfer, which results in inferior performance. Thus, a key challenge in minor language news recommendation lies in ensuring that news embeddings capture both content information and preference patterns.

Inspired by the observation that preference patterns are similar across languages (Schrøder, 2019; Guo et al., 2023), we propose a few-shot minor language news recommendation model based on cross-lingual **P**reference **P**attern **T**ransfer, named **PPT**. Specifically, PPT employs a Large Language Model (LLM)-based news encoder to learn accurate preference patterns in the many-shot English domain. Then, these patterns are transferred to the few-shot minor language domain by cross-lingual news augmentation. Additionally, through cross-lingual alignment, PPT extends the LLM-based news encoder's strong English encoding ability to minor languages for content representation, thus capturing both content and preference patterns.

Our contributions can be summarized as follows: (1) Based on the observation that preference patterns are learned by training, we identify the key challenge in minor language news recommendation as the difficulty of learning preference patterns due to limited interactions. (2) To address the challenge of interaction scarcity, we propose PPT, a minor language news recommendation model based on cross-lingual preference pattern transfer, which incorporates both news content and preference patterns. (3) We conduct extensive experiments on two real-world news recommendation datasets across a total of 15 minor languages. The consistent superior performance of PPT demonstrates its effectiveness and generalization.

## 2 Related Work

### 2.1 Basic English Recommendation

**Deep Neural-based Models**. With the development of deep learning, many deep neural-based methods have been proposed, such as NRMS (Wu et al., 2019b), NAML (Wu et al., 2019a), and LSTUR (An et al., 2019). These methods follow the two-tower architecture, employing deep neural networks as news and user encoders. However, limited by the size of networks, they struggle to fully capture the semantic information.

**LLM-based Models**. Given LLMs' powerful text comprehension capabilities, recent studies have explored their use as encoders. Models like NoteLLM (Zhang et al., 2024) and ONCE (Liu et al., 2024) enrich news content using LLMs and encode it with LLMs. While models like KAR (Xi et al., 2024) and LLMRec (Wei et al., 2024) utilize LLMs to encode the textual descriptions of both items and users to obtain embeddings. However, since most LLMs are primarily pre-trained on English corpora, they excel at encoding English but struggle with minor languages (Qin et al., 2025). Consequently, directly employing LLMs as encoders for minor language news recommendation results in unsatisfactory performance.

### 2.2 Cross-Lingual Recommendation

**Cold-Start Models**. Minor language recommendation can be considered a special case of cold-start recommendation (Narducci et al., 2016). Inspired by cross-domain recommender systems like CATN (Zhao et al., 2020) and CDRIB (Cao et al., 2022), the cold-start problem can be mitigated by bridging two domains through overlapping users. However, in the context of cross-lingual news recommendation, overlapping users are almost non-existent (Banks, 2011). A few-shot cross-lingual recommender system by sharing encoders does not rely on overlapping users (Guo et al., 2023). However, this method was proposed before the emergence of LLMs and, therefore, failed to leverage LLMs' strong English encoding capabilities. Additionally, it did not analyze the process of preference pattern transfer and is limited to the study between English and Norwegian only, thus lacking generalization.

**Multilingual Pre-trained Language Models** (mPLMs). Using mPLMs is a natural approach for cross-lingual recommendation. xMIND is a multilingual news recommendation dataset introduced in (Iana et al., 2024), where the authors encoded news articles by an mPLM, XLM-RoBERTa (Conneau, 2019). However, while mPLMs help reduce the language gap, their small scale limits their encoding capabilities, and they struggle with languages excluded from pre-training corpora.

In contrast, our proposed PPT leverages an LLM-based news encoder to capture content information and extend its strong encoding capability in English to minor languages by cross-lingual alignment. With cross-lingual news augmentation, PPT further transfers preference patterns from the English domain to the minor language domain, thus integrating content and preference patterns to improve minor language recommendation performance.

## 3 Methodology

The overview of PPT is illustrated in Figure 2. We adopt the widely used two-tower architecture. Both the news encoder and user encoder are shared and jointly trained in the English and minor language domains. This design enables effective preference pattern transfer through cross-lingual news augmentation and alignment.

### 3.1 Problem Definition

We aim to enhance minor language news recommendation by leveraging the rich interactions in the English domain. The problem studied in this paper is formulated as follows. Let $E$ denote the English domain and $M$ the minor language domain. The user and news sets in $E$ and $M$ are represented as $\mathcal{U}^E$,$\mathcal{U}^M$ and $\mathcal{D}^E$, $\mathcal{D}^M$, respectively. Each news $d \in \mathcal{D}$ is represented as a token sequence of its title, denoted as $[w_1, ..., w_{len(d)}]$. For a given user $u \in \mathcal{U}$, the set of clicked news history is denoted as $H_u = \{d_1, ..., d_{len(u)}\}$. Our goal is to predict the probability that a user $u^M$ in the minor language domain will click on a candidate news $d_c^M$.

### 3.2 Base Recommendation Model

#### 3.2.1 News Encoder

Considering the great success of LLMs in natural language processing, particularly in English tasks, we develop an LLM-based news encoder, which comprises an LLM, a linear layer, and an additive attention layer. Through LLM's inherent strong encoding capability and the training on interactions, the news encoder effectively captures both news content and preference patterns. LLM in PPT defaults to LLaMA-2-7b (Touvron et al., 2023).

**LLM** maps the token sequence of news $d$ into a continuous embedding space, and obtains the output embedding sequence $E^o \in \mathbb{R}^{len(d) \times k^l}$ through multiple Transformer layers, where $k^l$ is the dimension of LLM's embedding space. Given the strong inherent text comprehension capability of LLMs, all LLM parameters are frozen, and more details about fine-tuning can be found in Appendix B.3.

**Linear Layer** is trained to map $E^o$ from the LLM's large $k^l$-dimensional space to a much smaller $k^s$-dimensional space, denoted as $Z \in \mathbb{R}^{len(d) \times k^s}$. Thus reducing both storage requirements and computational costs.

**Additive Attention Layer** assigns different weights to each token and calculates the weighted sum as the final news embedding $e_d \in \mathbb{R}^{k^s}$:

$$a_i = q^T \tanh(W \times Z_i + b), \tag{1}$$

$$\alpha_i = \frac{\exp(a_i)}{\sum_{j=1}^{len(d)} \exp(a_j)}, \tag{2}$$

$$e_d = \sum_{i=1}^{len(d)} \alpha_i Z_i, \tag{3}$$

where $\alpha_i$ is the attention score of the $i_{th}$ token, while $q$, $W$ and $b$ are the trainable parameters.

#### 3.2.2 User Encoder

The user embedding is inferred from the clicked news history $H_u$. Since different clicked news articles contribute unequally to user representation, we use the same computational method as the additive attention layer in News Encoder to derive a more informative user embedding by emphasizing the selection of significant clicked news embeddings. The user embedding $e_u$ is defined by:

$$e_u = Attention(NewsEncoder(H_u)) \tag{4}$$

#### 3.2.3 Training and Inference

The click probability score is computed as the inner product of the user embedding and the candidate news embedding, $\hat{y}_{ud_c} = e_u^T e_{d_c}$. During training, we employ negative sampling techniques and set the number of negative samples

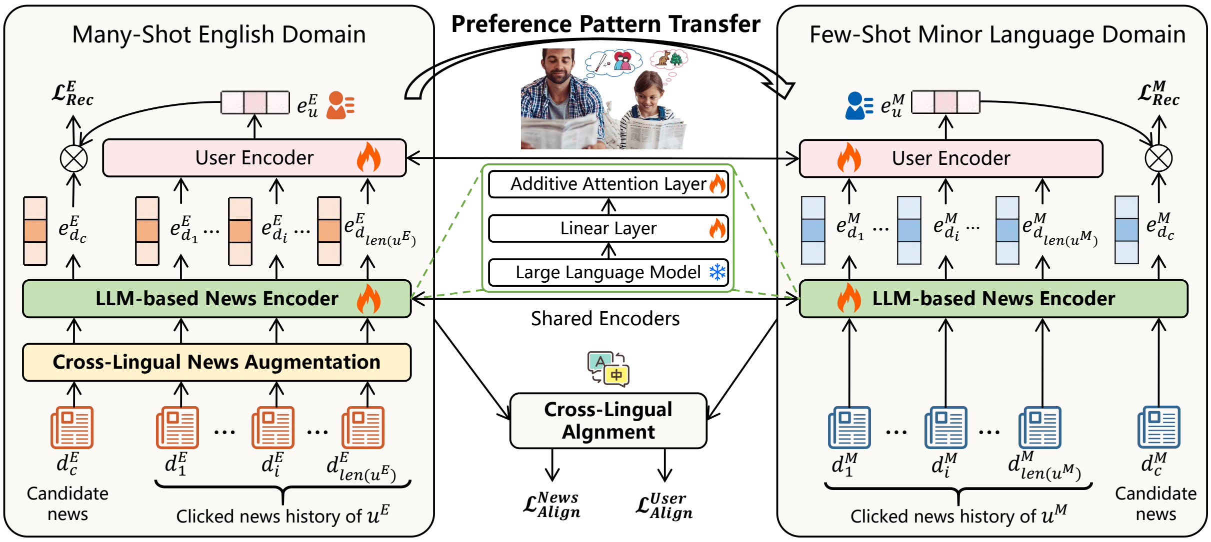



Figure 2: An overview of our proposed PPT.

to 4. Denote the total five candidate news as $[d^+, d_1^-, d_2^-, d_3^-, d_4^-]$, and their corresponding prediction scores as $[\hat{y}_{ud^+}, \hat{y}_{ud_1^-}, \hat{y}_{ud_2^-}, \hat{y}_{ud_3^-}, \hat{y}_{ud_4^-}]$. The recommendation loss function $L_{Rec}$ is the Cross-Entropy Loss of all positive samples $S$:

$$p_{ud^+} = \frac{\exp(\hat{y}_{ud^+})}{\exp(\hat{y}_{ud^+}) + \sum_{j=1}^{4} \exp(\hat{y}_{ud_j^-})}, \tag{5}$$

$$\mathcal{L}_{Rec} = \sum_{d^+ \in S} -\log(p_{ud^+}) \tag{6}$$

### 3.3 Cross-Lingual News Augmentation

The gap between embeddings of different languages hinders preference pattern transfer. While mPLMs can partially reduce this gap, their small scale prevents them from matching the encoding performance of state-of-the-art LLMs, especially in English. To address this, we propose cross-lingual news augmentation to leverage the LLM-based news encoder's strong English encoding capability and facilitate preference pattern transfer.

First, we translate each English news $d^E$ into its minor language counterpart $d^{TM}$ by the open-source NLLB-3.3B (Costa-jussà et al., 2022) machine translation system:

$$d^{TM} = NLLB(d^E), \tag{7}$$

thus, each news in the English domain is augmented into a set $A_d = \{d^E, d^{TM}\}$. Then, we randomly select one element from $A_d$ for news encoding during the training phase. When $d^{TM}$ is selected, it simulates interactions between minor language news and users in the English domain, which enables the transfer of preference patterns from the English domain to the minor language domain. Notably, user embeddings will not be distorted since $d^{TM}$ originates from $d^E$ and has similar semantic information. The subsequent cross-lingual alignment further ensures this.

### 3.4 Cross-Lingual Alignment

News alignment for $A_d$ helps bridge the gap between languages and extend the LLM's strong encoding capability from English to minor language. Instead of using the MSE loss as in (Guo et al., 2023), we use the Normalized Temperature-scaled Cross Entropy Loss (NT-Xent Loss) (Chen et al., 2020) for news alignment. By pulling positive pairs (news in the same $A_d$) closer and pushing negative pairs apart, NT-Xent Loss can enhance cross-lingual embedding consistency and improve generalization. The loss is formulated as:

$$p_i = \frac{\exp(sim(e_{d_i}^E, e_{d_i}^{TM})/\tau)}{\sum_{j=1, j\neq i}^{N} \sum_{e_{d_j} \in A_{d_j}} \exp(sim(e_{d_i}^E, e_{d_j})/\tau)}, \tag{8}$$

$$\mathcal{L}_{Align}^{News} = \frac{1}{N} \sum_{i=1}^{N} -\log p_i, \tag{9}$$

where $N$ is the number of news sets in one batch, and $\tau$ is the temperature hyperparameter. $sim$ denotes the cosine similarity. By minimizing $\mathcal{L}_{Align}^{News}$ to align news embeddings in different languages, we improve the LLM-based news encoder's ability to encode minor language news and mitigate the impact of translation quality for encoding $d^{TM}$.

Similarly, we also introduce the NT-Xent Loss $\mathcal{L}^{User}_{Align}$ to align user embeddings $e^{E}_{u}$ and $e^{TM}_{u}$, thus further ensuring cross-lingual alignment.

The overall loss function consists of the recommendation losses and the alignment losses:

$$\mathcal{L} = \alpha\mathcal{L}^{E}_{Rec} + \beta\mathcal{L}^{M}_{Rec} + \gamma\mathcal{L}^{News}_{Align} + \sigma\mathcal{L}^{User}_{Align} \quad (10)$$

where $\alpha$, $\beta$, $\gamma$, and $\sigma$ are the trade off hyperparameters, and Appendix B.2 provides the analysis.

## 4 Experiments

### 4.1 Experimental Setting

**Datasets**. We employ the widely used news recommendation dataset MIND(Wu et al., 2020) from Microsoft News[1] as the many-shot English domain and evaluate on two minor language news recommendation datasets xMIND(Iana et al., 2024) and Adressa(Gulla et al., 2017). xMIND, derived from MIND, covers 14 minor languages (see Table 1), with further details provided in Appendix A. To prevent user overlap and simulate the scenario of limited interactions on minor language news platforms, we include only warm-start users in MIND and cold-start users in xMIND. Warm-start users are defined as those with more than five clicked news, while the rest are classified as cold-start users. Adressa is a Norwegian news recommendations dataset collected from Adresseavisen[2], a news website in Norway. Detailed statistics for all experimental datasets are shown in Table 2. Our goal is to enhance the minor language news recommendation by transferring preference patterns from the many-shot English domain (MIND) to the few-shot minor language domains (xMIND and Adressa), thus assisting non-English-speaking countries in developing their own online news platforms.

| Code | Language | Code | Language |
|---|---|---|---|
| SWH | Swahili | THA | Thai |
| SOM | Somali | RON | Romanian |
| CMN | Traditional Chinese | FIN | Finnish |
| JPN | Japanese | KAT | Georgian |
| TUR | Turkish | HAT | Haitian Creole |
| TAM | Tamil | IND | Indonesian |
| VIE | Vietnamese | GRN | Guarani |

Table 1: The 14 languages included in xMIND.

[1]https://news.microsoft.com/source/

[2]https://www.adressa.no/

| | MIND | xMIND | Adressa |
|---|---|---|---|
| News | 50,323 | 24,195 | 29,799 |
| Users | 37,021 | 22,232 | 19,640 |
| Impressions | 130,019 | 39,928 | 19,640 |
| Avg. History | 38.66 | 2.96 | 8.71 |

Table 2: Statistics of experimental dataset.

**Evaluation Metrics**. Following state-of-the-art methods, we evaluate performance using widely adopted ranking metrics, including UAUC (User Area Under the Curve) (Zhou et al., 2018), MRR (Mean Reciprocal Rank) (Voorhees et al., 1999), and nDCG@K (normalized Discounted Cumulative Gain), where K is set to 1, 5, and 10, respectively (Järvelin and Kekäläinen, 2002).

**Baselines**. To evaluate performance, we compare PPT against several variants of DIRE (Liu et al., 2024) and a few-shot cross-lingual news recommendation method, Cross-Lingual-NRMS (Guo et al., 2023). In our experiments, DIRE-LLaMA represents DIRE with LLaMA-2-7b (Touvron et al., 2023) as the news encoder, while DIRE-XLM-RoBERTa (Conneau, 2019) represents using XLM-RoBERTa-base. ENG denotes the English dataset MIND with warm users, and ML denotes the Minor Language dataset with cold users. Following the Translate-then-align method in (Zhang et al., 2021), we further enhance some baselines through machine translation, denoted as Trans. .

- **DIRE-XLM-RoBERTa (ML+ML)** (Liu et al., 2024), using XLM-RoBERTa-base as the news encoder. Trained on the Minor Language dataset and tested on the Minor Language dataset.
- **DIRE-LLaMA (ML+ML)** (Liu et al., 2024), using LLaMA-2-7b as the news encoder. Trained on the Minor Language dataset and tested on the Minor Language dataset.
- **DIRE-LLaMA (Trans. ML+Trans. ML)** (Liu et al., 2024), using LLaMA-2-7b as the news encoder. Trained on the Minor Language dataset translated into English and tested on the Minor Language dataset translated into English.
- **DIRE-XLM-RoBERTa (ENG+ML)** (Liu et al., 2024), using XLM-RoBERTa-base as the news encoder. Trained on the English dataset and tested on the Minor Language dataset.
- **DIRE-LLaMA (ENG+ML)** (Liu et al., 2024), using LLaMA-2-7b as the news encoder. Trained on the English dataset and tested on the

Minor Language dataset.

- **DIRE-LLaMA (ENG+Trans. ML)** (Liu et al., 2024), using LLaMA-2-7b as the news encoder. Trained on the English dataset and tested on the Minor Language dataset translated into English.
- **Cross-Lingual-NRMS** (Guo et al., 2023), a few-shot news recommendation method by sharing encoders, which uses NRMS (Wu et al., 2019b) as the base recommendation model.

**Implementation Details**. We use LLaMA-2-7b (Touvron et al., 2023) as the default backbone. All experiments are conducted on a Linux server equipped with eight NVIDIA Tesla A40s. PPT is trained using the Adam optimizer with a learning rate of 1e-5. The maximum number of training epochs is set to 20, with an early stopping patience of 5. The hyperparameter $\tau$ in $\mathcal{L}_{Align}^{News}$ and $\mathcal{L}_{Align}^{User}$ is set as 0.05. The trade off hyperparameters of the overall loss function $\mathcal{L}$ in Equation 10 are set as $\alpha = 3$, $\beta = 1$, $\gamma = 1$, and $\sigma = 1$, respectively. The small embedding dimension $k^s$ after the linear layer of the new encoder is set to 64. We repeat each experiment five times with different random seeds and report the mean results.

### 4.2 Overall Performance

Table 3 presents the average recommendation performance of PPT and baselines on xMIND across 14 minor languages. Performance for each minor language is provided in Appendix B.4. From the results, we can draw the following findings:

- On average, PPT significantly outperforms all baselines across five metrics, with gains of at least 3.38%, 6.85%, 12.48%, 7.70%, and 5.55%. The strong English comprehension capability of the LLM-based news encoder enables PPT to learn accurate preference patterns in the many-shot English domain. Cross-lingual news augmentation transfers these patterns to the few-shot minor language domain, while cross-lingual alignment extends the encoder's English encoding ability to minor languages for content understanding. As a result, PPT effectively captures both content information and preference patterns, thus achieving remarkable minor language news recommendation performance.
- Among the baselines, DIRE-LLaMA (Trans. ML+Trans. ML) and DIRE-LLaMA (ENG +Trans.ML) perform relatively better. Both models are trained and tested on English news, thus fully leveraging LLM's strong English encoding capability. However, DIRE-LLaMA (Trans.ML+Trans.ML) is trained in the few-shot domain, so it fails to fully learn preference patterns. On the other hand, although DIRE-LLaMA (ENG+Trans. ML) is trained in the many-shot English domain, semantic shifts caused by imperfect translations can undermine the modeling of content information. As a result, both translation-based models still suffer from ineffective preference pattern transfer or inferior news content representation.
- DIRE-LLaMA (ENG+ML) performs poorly, which suggests that training in the many-shot English domain and inferring in the minor language domain is ineffective for preference pattern transfer. This can be attributed to the gap between English and minor languages in the LLM's embedding space. As for DIRE-XLM-RoBERTa (ML+ML) and DIRE-XLM-RoBERTa (ENG+ML), which use the multilingual language model XLM-RoBERTa-base as the news encoder, their performance remains unsatisfactory due to the inferior encoding capability of mPLMs compared to LLMs and the ineffective preference pattern transfer.

To further analyze the performance across xMIND's 14 minor languages individually, we present the nDCG@10 results for all languages in Figure 3. As expected, PPT achieves the best performance across most languages, except for KAT, where PPT ranks second to DIRE-LLaMA (ENG+Trans.ML). This can be attributed to KAT being a low-resource language with only 3.9 million speakers (Iana et al., 2024) and using the unique Georgian script rather than the Latin script adopted by English and most other languages. This results in a larger gap between KAT and ENG, which makes KAT news recommendation more challenging. Among the baselines, the translation-based DIRE-LLaMA models, which benefit from LLM's strong English encoding capability, perform well in most languages. However, their performance is limited by translation quality (see Appendix B.8). For instance, they perform poorly in CMN, whose translation quality is low (BLEU=0.1127).

Considering that xMIND and MIND both originate from Microsoft News, we further evaluate PPT on a more challenging dataset, Adressa, which is from the Norwegian online news plat-

| Model | UAUC | MRR | nDCG@1 | nDCG@5 | nDCG@10 |
|---|---|---|---|---|---|
| DIRE-XLM-RoBERTa (ML+ML) | 56.95 | 29.70 | 14.47 | 30.51 | 36.80 |
| DIRE-LLaMA (ML+ML) | 55.36 | 29.20 | 14.44 | 29.97 | 36.08 |
| DIRE-LLaMA (Trans. ML+Trans. ML) | <u>57.45</u> | 30.64 | <u>15.78</u> | 31.65 | 37.62 |
| DIRE-XLM-RoBERTa (ENG+ML) | 55.42 | 29.52 | 14.70 | 30.07 | 36.39 |
| DIRE-LLaMA (ENG+ML) | 49.56 | 26.18 | 11.49 | 26.32 | 32.43 |
| DIRE-LLaMA (ENG+Trans. ML) | 56.94 | <u>30.66</u> | 15.52 | <u>31.67</u> | 37.71 |
| Cross-Lingual-NRMS | 56.56 | 30.55 | 15.26 | 31.55 | <u>37.86</u> |
| PPT | **59.39** | **32.76** | **17.75** | **34.11** | **39.96** |
| Improvement over the best baseline | 3.38% | 6.85% | 12.48% | 7.70% | 5.55% |

Table 3: Average performance comparison on **xMIND**. The best results are highlighted in **bold**, and the second-best results are <u>underlined</u>.

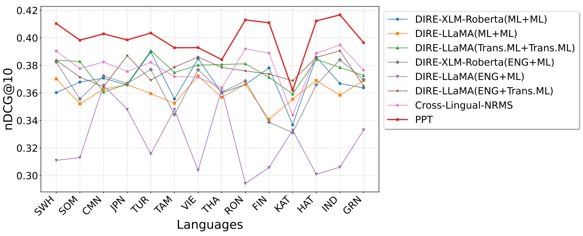
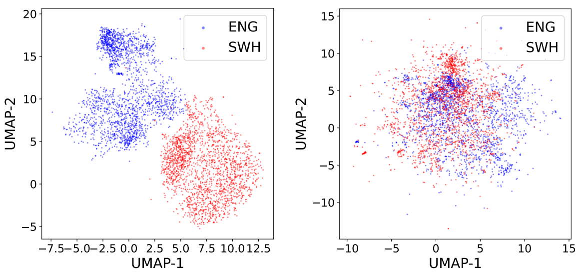



Figure 3: Performance comparison across all languages on xMIND.




(a) Initial embeddings

(b) Trained embeddings

Figure 4: The visualization of English and minor language (take SWH of xMIND as example) news embeddings before and after training.

form Adresseavisen. As shown in Table 4, PPT still performs best across all metrics. This demonstrates PPT's ability to transfer preference patterns in more practical scenarios (from Microsoft News to Adresseavisen). The consistent superior performance also shows PPT's generalization.

To validate that PPT effectively bridges the gap between English and minor languages by preference pattern transfer, we apply UMAP (McInnes et al., 2018) to project the news embeddings from both English and the minor language before and after training into a two-dimensional space, as visualized in Figure 4. Before training, the initial English and SWH news embeddings are two separate clusters, which exhibits a clear gap. After training, the news embeddings of the two languages become closer and intermixed. This indicates that our proposed cross-lingual strategies effectively reduce the lingual shift and facilitate the transfer of preference patterns from the English domain to the minor language domain.

To verify the effectiveness of each loss, we analyze the loss trends over training steps, as illustrated in Figure 5. The results show that as the training progresses, all losses show a decreasing trend and gradually converge, demonstrating that PPT effectively optimizes each loss function.

The above experiments focus on few-shot recommendation. To further validate PPT's generalization, we also evaluate its performance under a more challenging zero-shot setting ($\beta$ set to 0). The experimental results are provided in Appendix B.6.

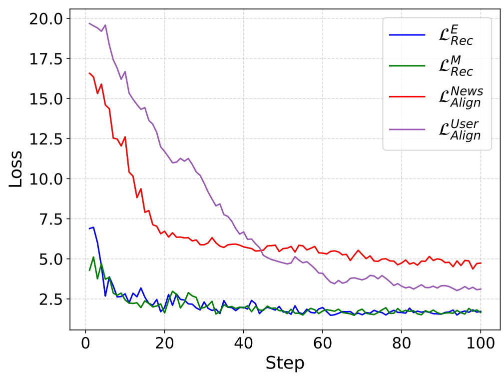



Figure 5: The loss trend.

### 4.3 Ablation Studies

In this section, we examine the impact of cross-lingual news augmentation and alignment. The average results on 14 minor languages of xMIND are presented in Table 5, and more detailed studies for each language can be found in Appendix B.5. The following observations can be drawn:

| Model | UAUC | MRR | nDCG@1 | nDCG@5 | nDCG@10 |
|---|---|---|---|---|---|
| DIRE-XLM-RoBERTa (ML+ML) | 59.06 | 64.11 | 58.39 | 60.97 | 69.64 |
| DIRE-LLaMA (ML+ML) | 58.32 | 63.69 | 59.15 | 60.72 | 70.05 |
| DIRE-LLaMA (Trans. ML+Trans. ML) | 58.65 | 64.00 | <u>59.28</u> | 60.96 | <u>70.29</u> |
| DIRE-XLM-RoBERTa (ENG+ML) | 58.08 | 63.62 | 59.04 | 59.88 | 69.73 |
| DIRE-LLaMA (ENG+ML) | 52.89 | 61.42 | 50.86 | 56.76 | 67.38 |
| DIRE-LLaMA (ENG+Trans. ML) | 57.75 | 63.40 | 58.07 | 59.82 | 69.64 |
| Cross-Lingual-NRMS | <u>59.18</u> | <u>64.15</u> | 58.72 | <u>61.00</u> | 69.97 |
| PPT | **60.37** | **65.78** | **60.05** | **62.24** | **71.64** |
| Improvement over the best baseline | 2.01% | 2.54% | 1.30% | 2.03% | 1.92% |

Table 4: Performance comparison on **Adressa**. The best results are highlighted in **bold**, and the second-best results are <u>underlined</u>.

| w/o | UAUC | MRR | nDCG@1 | nDCG@5 | nDCG@10 |
|---|---|---|---|---|---|
| / | **59.39** | **32.76** | **17.75** | **34.11** | **39.96** |
| Cross-Lingual News Augmentation | 59.19 | 32.61 | 17.67 | 33.94 | 39.76 |
| Cross-Lingual Alignment | 58.47 | 32.04 | 17.17 | 33.30 | 39.09 |
| Both Augmentation & Alignment | 58.01 | 31.36 | 16.64 | 32.43 | 38.40 |

Table 5: The impact of cross-lingual news augmentation and alignment. *w/o* denotes without.

- The full PPT achieves the best performance across all metrics. In the "w/o news augmentation" and "w/o alignment" settings, the performance declines, which indicates that both cross-lingual strategies contribute positively. Additionally, they complement each other in mitigating the lingual shift and transferring preference patterns from the many-shot English domain to the few-shot minor language domain.

- The "w/o both augmentation& alignment" setting yields the worst performance, as it only involves training in the English and minor language domains without any cross-lingual enhancements. Consequently, there remains a significant gap between English and minor languages, which prevents effective preference pattern transfer and accurate news content representation.

To further validate the effectiveness of cross-lingual news augmentation and alignment, Figure 6 presents the percentage improvement of PPT over various w/o settings across 14 minor languages in xMIND. For most minor languages, either cross-lingual news augmentation or alignment can improve performance. Additionally, incorporating both cross-lingual augmentation & alignment consistently improves performance across all 14 minor languages, highlighting the effectiveness of our cross-lingual strategies in reducing the lingual shift and transferring preference patterns.

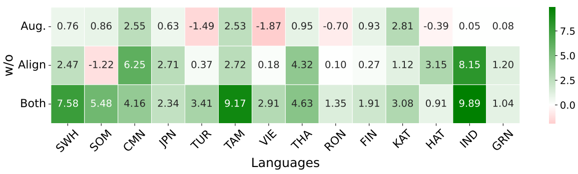



Figure 6: Relative percentage improvements of complete PPT over various w/o settings. *Aug.* denotes cross-lingual news augmentation, *Align* denotes cross-lingual Alignment, and *Both* denotes both *Aug.* and *Align*.

### 4.4 Experiments with the Language Models

PPT defaults to using LLaMA-2-7b as the backbone for the LLM-based news encoder. However, PPT is compatible with any other language model. In this section, we validate the advantages of using LLMs like LLaMA-2-7b over other traditional PLMs such as BERT-base and even mPLMs like XLM-RoBERTa-base.

As shown in Table 6, PPT with LLaMA-2-7b achieves significantly better results. More detailed results are provided in Appendix B.7. While LLaMA's multilingual capability is weaker than XLM-Roberta due to its primary training on English texts (Engländer et al., 2024), LLaMA's English encoding capability is stronger because of its larger parameter size and training on extensive English corpora. By cross-lingual augmentation and alignment, LLaMA's strong English encoding capability is extended to minor languages, thus achieving remarkable performance. These findings highlight the potential of LLMs in developing

| Pre-trained Model | UAUC | MRR | nDCG@1 | nDCG@5 | nDCG@10 |
|---|---|---|---|---|---|
| LLaMA-2-7b | **59.39** | **32.76** | **17.75** | **34.11** | **39.96** |
| BERT-base | 55.07 | 29.56 | 14.03 | 30.39 | 36.57 |
| XLM-RoBERTa-base | 57.02 | 30.44 | 15.03 | 31.52 | 37.83 |

Table 6: Average performance with different pre-trained model.

online news platforms for non-English-speaking countries and even advancing other minor language tasks. This ensures that LLMs can benefit a global audience rather than just English speakers.

## 5 Conclusion

Minor language online news platforms struggle to learn preference patterns due to insufficient user-news interactions, which makes the personalized recommendation challenging. To solve this problem, we propose a minor language news recommendation model based on cross-lingual preference pattern transfer, named PPT. Specifically, we develop an LLM-based news encoder and extend its strong English encoding capability to minor languages via cross-lingual alignment. Additionally, we simulate interactions of minor language news in the English domain by cross-lingual news augmentation, thus further facilitating preference pattern transfer. Extensive experimental results on two real-world datasets across 15 minor languages show that PPT consistently outperforms the baselines, which demonstrates PPT's superiority and generalization.

## 6 Limitations

The main limitations of this paper are as follows: (1) Due to limited computational resources, we do not fine-tune the LLM of the news encoder. Various LLM fine-tuning strategies like LoRA could be further tested to assess their effectiveness in helping PPT better learn and transfer preference patterns. (2) While we conducted extensive offline experiments on a total of 15 languages across two minor language datasets, real-world online validation remains lacking. In future work, we plan to fine-tune LLMs and seek opportunities for practical online A/B testing to further demonstrate PPT's potential in improving minor language recommendation performance through preference pattern transfer.

## Acknowledgments

We thank anonymous reviewers for their insightful feedback. We also thank Joey Xu at Columbia University for her help in English writing.

## References

Mingxiao An, Fangzhao Wu, Chuhan Wu, Kun Zhang, Zheng Liu, and Xing Xie. 2019. Neural news recommendation with long-and short-term user representations. In *Proceedings of the 57th annual meeting of the association for computational linguistics*, pages 336–345.

Matthew Banks. 2011. *The Same, but Different: An Exploratory Study of English as Second Language Learners in the Context of their Information Behaviour Related to Recreational Reading*. Ph.D. thesis, Open Access Te Herenga Waka-Victoria University of Wellington.

Jiangxia Cao, Jiawei Sheng, Xin Cong, Tingwen Liu, and Bin Wang. 2022. Cross-domain recommendation to cold-start users via variational information bottleneck. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, pages 2209–2223. IEEE.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR.

A Conneau. 2019. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*.

Marta R Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, et al. 2022. No language left behind: Scaling human-centered machine translation. *arXiv preprint arXiv:2207.04672*.

Don Cusic. 2003. *Baseball and country music*. Popular Press.

Leon Engländer, Hannah Sterz, Clifton Poth, Jonas Pfeiffer, Ilia Kuznetsov, and Iryna Gurevych. 2024. M2qa: Multi-domain multilingual question answering. *arXiv preprint arXiv:2407.01091*.

Jon Atle Gulla, Lemei Zhang, Peng Liu, Özlem Özgöbek, and Xiaomeng Su. 2017. The adressa dataset for news recommendation. In *Proceedings of the international conference on web intelligence*, pages 1042–1048.

Taicheng Guo, Lu Yu, Basem Shihada, and Xiangliang Zhang. 2023. Few-shot news recommendation via cross-lingual transfer. In *Proceedings of the ACM Web Conference 2023*, pages 1130–1140.

Andreea Iana, Goran Glavaš, and Heiko Paulheim. 2024. Mind your language: a multilingual dataset for cross-lingual news recommendation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 553–563.

Kalervo Järvelin and Jaana Kekäläinen. 2002. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4):422–446.

Qijiong Liu, Nuo Chen, Tetsuya Sakai, and Xiao-Ming Wu. 2024. Once: Boosting content-based recommendation with both open-and closed-source large language models. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, pages 452–461.

Leland McInnes, John Healy, and James Melville. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

Fedelucio Narducci, Pierpaolo Basile, Cataldo Musto, Pasquale Lops, Annalina Caputo, Marco de Gemmis, Leo Iaquinta, and Giovanni Semeraro. 2016. Concept-based item representations for a cross-lingual content-based recommendation process. *Information Sciences*, 374:15–31.

Libo Qin, Qiguang Chen, Yuhang Zhou, Zhi Chen, Yinghui Li, Lizi Liao, Min Li, Wanxiang Che, and S Yu Philip. 2025. A survey of multilingual large language models. *Patterns*, 6(1).

K Schrøder. 2019. What do news readers really want to read about? how relevance works for news audiences. *Digital News Project*, (Feb 2019).

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

Patricia Vignola. 2005. Baseball and country music. *NINE: A Journal of Baseball History and Culture*, 13(2):169–272.

Ellen M Voorhees et al. 1999. The trec-8 question answering track report. In *Trec*, volume 99, pages 77–82.

Wei Wei, Xubin Ren, Jiabin Tang, Qinyong Wang, Lixin Su, Suqi Cheng, Junfeng Wang, Dawei Yin, and Chao Huang. 2024. Llmrec: Large language models with graph augmentation for recommendation. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, pages 806–815.

Chuhan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang, and Xing Xie. 2019a. Neural news recommendation with attentive multi-view learning. *arXiv preprint arXiv:1907.05576*.

Chuhan Wu, Fangzhao Wu, Suyu Ge, Tao Qi, Yongfeng Huang, and Xing Xie. 2019b. Neural news recommendation with multi-head self-attention. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, pages 6389–6394.

Chuhan Wu, Fangzhao Wu, Yongfeng Huang, and Xing Xie. 2023. Personalized news recommendation: Methods and challenges. *ACM Transactions on Information Systems*, 41(1):1–50.

Fangzhao Wu, Ying Qiao, Jiun-Hung Chen, Chuhan Wu, Tao Qi, Jianxun Lian, Danyang Liu, Xing Xie, Jianfeng Gao, Winnie Wu, et al. 2020. Mind: A large-scale dataset for news recommendation. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 3597–3606.

Yunjia Xi, Weiwen Liu, Jianghao Lin, Xiaoling Cai, Hong Zhu, Jieming Zhu, Bo Chen, Ruiming Tang, Weinan Zhang, and Yong Yu. 2024. Towards open-world recommendation with knowledge augmentation from large language models. In *Proceedings of the 18th ACM Conference on Recommender Systems*, pages 12–22.

Chao Zhang, Shiwei Wu, Haoxin Zhang, Tong Xu, Yan Gao, Yao Hu, and Enhong Chen. 2024. Notellm: A retrievable large language model for note recommendation. In *Companion Proceedings of the ACM on Web Conference 2024*, pages 170–179.

Wenxuan Zhang, Ruidan He, Haiyun Peng, Lidong Bing, and Wai Lam. 2021. Cross-lingual aspect-based sentiment analysis with aspect term code-switching. In *Proceedings of the 2021 conference on empirical methods in natural language processing*, pages 9220–9230.

Cheng Zhao, Chenliang Li, Rong Xiao, Hongbo Deng, and Aixin Sun. 2020. Catn: Cross-domain recommendation for cold-start users via aspect transfer network. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*, pages 229–238.

Guorui Zhou, Xiaoqiang Zhu, Chenru Song, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, and Kun Gai. 2018. Deep interest network for click-through rate prediction. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1059–1068.

# A Details of xMIND

xMIND is a multilingual news recommendation dataset that covers 14 minor languages. Table 7 presents the detailed information of all languages. We conduct experiments across all 14 minor languages spanning six scripts and five macro-areas. The results demonstrate PPT's superiority and generalization.

| Code | Language | Script | Macro-area | Total speakers (M) |
|---|---|---|---|---|
| SWH | Swahili | Latin | Africa | 71.6 |
| SOM | Somali | Latin | Africa | 22.0 |
| CMN | Traditional Chinese | Han | Eurasia | 1,138.2 |
| JPN | Japanese | Japanese | Eurasia | 1,234.5 |
| TUR | Turkish | Latin | Eurasia | 90.0 |
| TAM | Tamil | Tamil | Eurasia | 86.6 |
| VIE | Vietnamese | Latin | Eurasia | 85.8 |
| THA | Thai | Thai | Eurasia | 60.8 |
| RON | Romanian | Latin | Eurasia | 24.5 |
| FIN | Finnish | Latin | Eurasia | 5.6 |
| KAT | Georgian | Georgian | Eurasia | 3.9 |
| HAT | Haitian Creole | Latin | North America | 13.0 |
| IND | Indonesian | Latin | Papunesia | 199.1 |
| GRN | Guarani | Latin | South America | 6.7 |

Table 7: Details of 14 languages of xMIND.

# B Additional Experimental Results

## B.1 The Relationship between Performance and Interactions.

As many studies have demonstrated, sufficient interactions are crucial for building recommender systems. To validate this, we evaluate on MIND using varying history sequence lengths. As shown in Table 8, the performance consistently declines as the history shortens, indicating the importance of interactions. A particularly sharp decline is observed when the length is reduced from 20 to 5, highlighting the difficulty of capturing accurate preference patterns under sparse interactions. Thus, building recommender systems for minor language news platforms with limited interactions remains a significant challenge.

| Lengths | UAUC | MRR | n@1 | n@5 | n@10 |
|---|---|---|---|---|---|
| 40 | 67.32 | 34.75 | 20.37 | 36.36 | 42.37 |
| 30 | 66.37 | 34.58 | 20.19 | 36.20 | 42.25 |
| 20 | 64.91 | 34.27 | 19.73 | 35.89 | 39.90 |
| 10 | 62.77 | 33.62 | 18.82 | 35.29 | 39.18 |
| 5 | 60.31 | 32.86 | 17.51 | 34.27 | 40.07 |

Table 8: Effect of the history sequence length.

## B.2 Hyperparameter Sensitivity Analysis.

The hyperparameters in the loss function $\mathcal{L}$ are used to balance the proportions of each loss component. In our paper, we set $\alpha = 3, \beta = 1, \gamma = 1, \sigma = 1$, which achieves satisfactory performance. This configuration fully leverages the rich user-news interactions in the English domain, effectively incorporating preference patterns into news representations. To study the influence of hyperparameters, Table 9 presents the hyperparameter sensitivity analysis. Results indicate that even with all hyperparameters set to 1, PPT still outperforms the best baseline. Actually, the ratio between $\alpha$ and $\beta$ has the greatest impact, where $\alpha$ is the weight of the recommendation loss in the English domain and $\beta$ is the weight of the recommendation loss in the minor language domain. A low ratio weakens preference patterns learning from the English domain, while a high ratio hinders adaptation to the minor language domain. Setting the ratio around 3 provides a good balance.

In conclusion, although hyperparameters do influence performance, setting $\alpha$ to approximately three times $\beta$ is sufficient to achieve strong performance. The results also demonstrate PPT's robustness to hyperparameters.

## B.3 Effect of Fine-tuning LLMs

As mentioned in the main text, considering the strong inherent text comprehension capability of LLMs, we do not fine-tune the LLM in our study. To further verify that fine-tuning is indeed unnecessary, we also experimented with using LoRA to fine-tune the last layer of the LLM in our study. As shown in Table 10, fine-tuning the LLM does not improve performance significantly. This can be attributed to the nature of our task, cross-lingual news recommendation, which requires training in both English and minor language domains. Since LLMs are less effective at understanding minor language texts, fine-tuning them on minor language interactions may degrade their ability to capture content information. Therefore, PPT freezes the LLM to preserve its content encoding capabilities and apply a trainable linear layer and an additive attention layer to learn user preference patterns and transfer them from the English domain to the

| Hyperparameter | UAUC | MRR | n@1 | n@5 | n@10 |
|---|---|---|---|---|---|
| $\alpha = 1, \beta = 1, \gamma = 1, \sigma = 1$ | 58.17 | 31.58 | 16.83 | 32.68 | 38.72 |
| $\alpha = 2, \beta = 1, \gamma = 1, \sigma = 1$ | 58.91 | 32.17 | 17.32 | 33.57 | 39.50 |
| $\alpha = 3, \beta = 1, \gamma = 1, \sigma = 1$ | **59.39** | **32.76** | **17.75** | **34.11** | **39.96** |
| $\alpha = 4, \beta = 1, \gamma = 1, \sigma = 1$ | 59.17 | 32.60 | 17.43 | 33.81 | 39.76 |
| $\alpha = 5, \beta = 1, \gamma = 1, \sigma = 1$ | 58.93 | 32.29 | 17.33 | 33.62 | 39.58 |
| $\alpha = 3, \beta = 2, \gamma = 1, \sigma = 1$ | 58.75 | 32.06 | 17.17 | 33.60 | 39.44 |
| $\alpha = 3, \beta = 1, \gamma = 2, \sigma = 1$ | 59.57 | 32.80 | 17.83 | 33.99 | 39.87 |
| $\alpha = 3, \beta = 1, \gamma = 1, \sigma = 2$ | 59.28 | 32.55 | 17.68 | 33.85 | 39.83 |
| $\alpha = 3, \beta = 1, \gamma = 2, \sigma = 2$ | 59.11 | 32.23 | 17.44 | 33.73 | 39.72 |
| $\alpha = 3, \beta = 2, \gamma = 1, \sigma = 2$ | 58.70 | 31.91 | 17.00 | 32.97 | 39.04 |
| $\alpha = 3, \beta = 2, \gamma = 2, \sigma = 1$ | 58.66 | 31.97 | 17.02 | 33.11 | 39.15 |

Table 9: Hyperparameter sensitivity analysis. The setting adopted in our paper is highlighted in **bold**.

minor language domain through proposed cross-lingual strategies. This enables PPT to efficiently capture both news content and preference patterns for minor language news recommendation while conserving computing resources.

| F/T | UAUC | MRR | n@1 | n@5 | n@10 |
|---|---|---|---|---|---|
| 32/0 | 59.39 | 32.76 | 17.75 | 34.11 | 39.96 |
| 31/1 | 59.46 | 32.77 | 17.88 | 34.04 | 40.02 |

Table 10: Effect of fine-tuning the LLM. *F/T* denotes the number of Frozen/Tuning layers.

### B.4 Performance Comparison Across All 14 Minor Languages of xMIND

Table 16 presents the performance comparison across all 14 minor languages on xMIND. Additionally, to further illustrate PPT's robustness and superior performance across different minor languages, we also provide line charts comparing PPT with the baselines on various metrics in Figure 7 (the results of nDCG@10 have already been presented in Figure 3). The results indicate that PPT consistently achieves the best performance across nearly all metrics for all minor languages on xMIND. The only exception is KAT, where PPT does not perform optimally on any of the five metrics. As discussed in the main text, we attribute this to KAT being a low-resource language and using the specialized Georgian script rather than the Latin script widely adopted by English and most other languages. This makes it particularly challenging for LLMs, which are primarily trained on English corpora, to encode KAT texts, resulting in a significant gap between KAT and ENG and leading to PPT's inferior performance in this case.

### B.5 Ablation Study Across All 14 Minor Languages of xMIND

Table 17 shows the impact of cross-lingual augmentation and alignment on each individual minor language on xMIND. The results indicate that in most cases, both news augmentation and alignment can further improve performance. This demonstrates that these two strategies effectively reduce lingual shift and enable the transfer of preference patterns learned from the many-shot English domain to the few-shot minor language domain.

### B.6 Zero-Shot Performance

For newly established news platforms, a limited user base can make the more challenging zero-shot recommendation problem. We evaluate the performance of zero-shot minor language news recommendation on xMIND by setting the hyperparameter $\beta$ to 0 in PPT's loss function. In other words, the training set excludes any interaction data from the minor language domain.

As shown in Table 11, PPT's recommendation performance in the zero-shot setting is slightly lower than that in the few-shot setting on xMIND. For some languages, such as VIE and HAT, zero-shot recommendation performance even outperforms few-shot performance. We speculate that this is due to the lack of user-news interactions in minor language datasets, which makes historical clicks an unreliable reflection of user preferences.

Additionally, since Adressa and MIND originate from two entirely different news platforms, we also evaluate the zero-shot performance on Adressa to further validate the effectiveness of PPT in a more challenging real-world scenario. As shown in Table 12, the performance drop from few-shot to zero-shot is relatively small, and PPT with zero-shot still beats all few-shot baselines. This demonstrates that

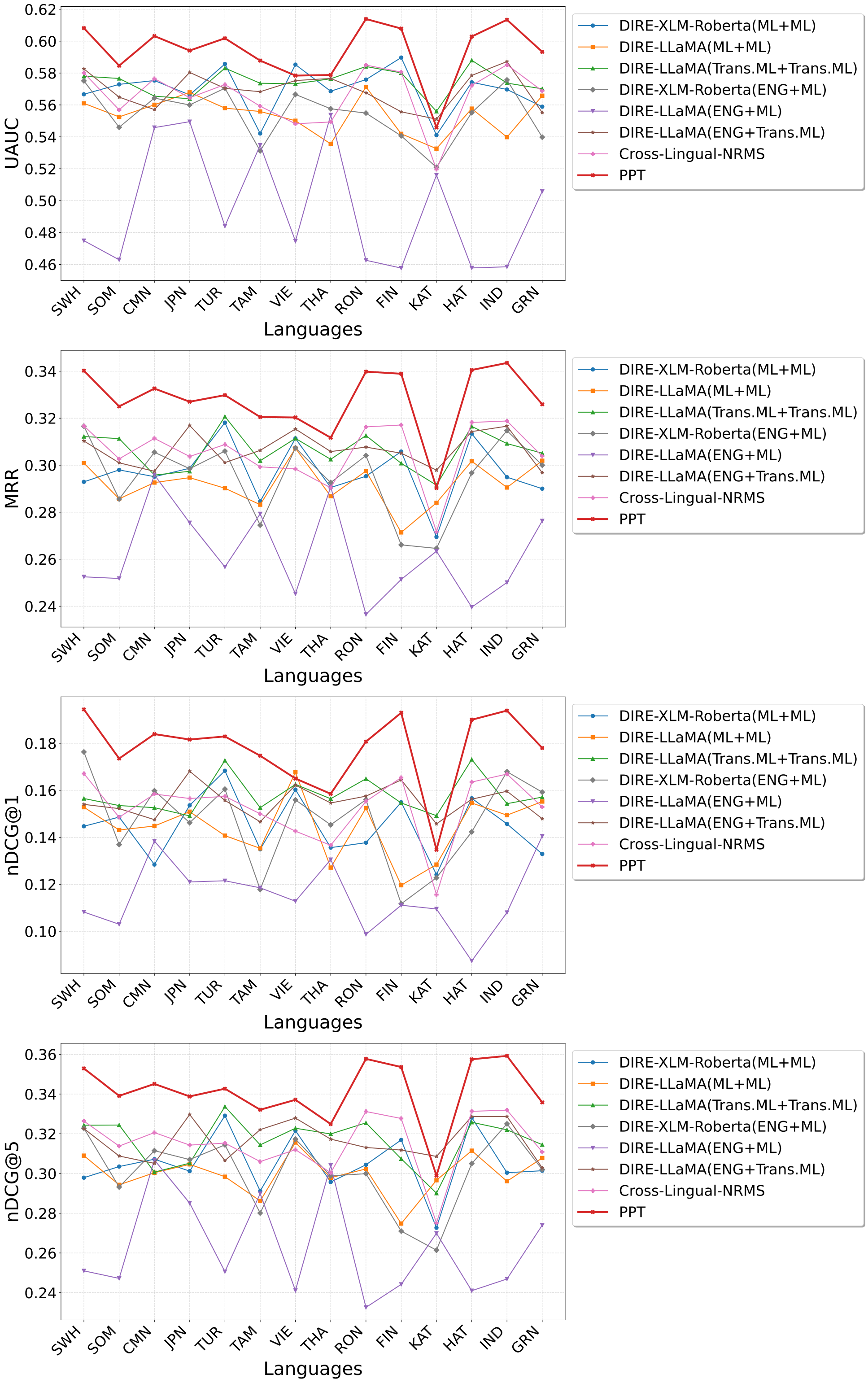



Figure 7: Performance across 14 minor languages.

| Language | Few-shot UAUC | MRR | n@1 | n@5 | n@10 | Zero-shot UAUC | MRR | n@1 | n@5 | n@10 |
|---|---|---|---|---|---|---|---|---|---|---|
| SWH | **60.82** | **34.02** | **19.44** | **35.29** | **41.04** | 60.44 | 33.49 | 18.38 | 35.09 | 40.71 |
| SOM | 58.46 | 32.50 | 17.35 | 33.91 | 39.83 | **58.59** | **32.63** | **17.46** | **34.02** | **39.83** |
| CMN | **60.32** | **33.26** | **18.39** | **34.51** | **40.29** | 59.44 | 32.36 | 17.37 | 33.40 | 39.49 |
| JPN | **59.41** | **32.70** | **18.16** | **33.88** | **39.85** | 57.92 | 31.26 | 16.20 | 32.39 | 38.63 |
| TUR | **60.18** | **32.98** | **18.29** | **34.27** | **40.35** | 59.65 | 32.34 | 16.72 | 33.91 | 39.80 |
| TAM | **58.78** | **32.05** | **17.47** | **33.21** | **39.28** | 55.40 | 30.28 | 15.59 | 31.15 | 37.10 |
| VIE | 57.84 | 32.03 | 16.50 | 33.71 | 39.29 | **58.83** | **32.96** | **18.06** | **34.65** | **40.13** |
| THA | **57.88** | **31.17** | **15.85** | **32.49** | **38.42** | 56.61 | 30.18 | 15.33 | 31.11 | 37.31 |
| RON | **61.39** | **33.98** | 18.07 | **35.77** | **41.30** | 60.72 | 33.85 | **18.33** | 35.61 | 41.05 |
| FIN | **60.79** | **33.89** | **19.30** | **35.36** | **41.10** | 59.67 | 32.72 | 17.76 | 33.93 | 40.05 |
| KAT | **54.59** | **29.03** | **13.47** | **29.89** | **36.19** | 52.47 | 27.80 | 12.56 | 28.49 | 34.55 |
| HAT | 60.29 | 34.05 | 19.00 | 35.75 | 41.23 | **60.76** | **34.49** | **19.19** | **36.07** | **41.50** |
| IND | **61.34** | **34.35** | **19.39** | **35.92** | **41.67** | 61.13 | 33.56 | 18.04 | 35.21 | 40.99 |
| GRN | **59.33** | 32.59 | 17.80 | 33.58 | 39.65 | 59.01 | **32.84** | **18.11** | **34.14** | **39.89** |
| Avg. | **59.39** | **32.76** | **17.75** | **34.11** | **39.96** | 58.62 | 32.20 | 17.08 | 33.51 | 39.36 |

Table 11: Performance comparison between few-shot and zero-shot on xMIND. The better results are highlighted in **bold**.

PPT can achieve strong performance even without any user-news interactions from the minor language domain during training, highlighting its real-world applicability and generalization capability.

| Shot | UAUC | MRR | n@1 | n@5 | n@10 |
|---|---|---|---|---|---|
| Few | **60.37** | **65.78** | **60.05** | **62.24** | **71.64** |
| Zero | 59.97 | 65.06 | 59.46 | 61.56 | 71.05 |

Table 12: Performance comparison between few-shot and zero-shot on Adressa. The better results are highlighted in **bold**.

## B.7 Effect of Pre-trained Language Models

Table 18 shows the impact of different PLMs on each minor language on xMIND. LLaMA-2-7b (the default PLM in our paper) consistently achieves the best performance, which demonstrates that PPT effectively bridges the gap between ENG and minor language while transferring preference patterns from the English domain to the minor language domain. Although LLaMA-2-7b is primarily designed for English tasks, its minor language comprehension capability can be enhanced by our proposed cross-lingual strategies, thus highlighting the immense potential of LLMs for minor language applications.

To further verify the effectiveness of our proposed cross-lingual strategies, we conduct the ablation study with XLM-RoBERTa-base and BERT-base on xMIND, as shown in Table 13. The results demonstrate that cross-lingual augmentation and alignment do achieve effective performance improvements with different LLM-based news encoders, which highlights PPT's generalization.

| ***BERT-base*** | | | | | |
|---|---|---|---|---|---|
| **w/o** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** |
| / | 55.07 | 29.56 | 14.03 | 30.39 | 36.57 |
| Aug. | 54.77 | 29.35 | 13.89 | 30.12 | 36.23 |
| Align. | 54.12 | 28.80 | 13.50 | 29.68 | 35.62 |
| Both | 53.60 | 28.21 | 13.15 | 29.01 | 35.20 |
| ***XLM-RoBERTa-base*** | | | | | |
| **w/o** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** |
| / | 57.02 | 30.44 | 15.03 | 31.52 | 37.83 |
| Aug. | 56.73 | 30.22 | 14.82 | 31.20 | 37.49 |
| Align. | 55.98 | 29.67 | 14.41 | 30.74 | 36.88 |
| Both | 55.58 | 29.23 | 14.11 | 30.06 | 36.36 |

Table 13: Ablation Study with different backbone PLMs on xMIND. The best results are highlighted in bold. *Aug.* denotes cross-lingual augmentation, and *Align.* denotes cross-lingual alignment. *w/o* denotes without.

## B.8 Effect of Translation Quality

Some of the baselines used for performance comparison require translating minor languages into English. Therefore, we computed the BLEU metric between English news titles translated from xMIND and the original English news titles in MIND to analyze the impact of translation quality on recommendation performance.

As shown in Table 14, minor languages that use the Latin script, such as HAT, SWH, and IND, generally exhibit higher translation quality, whereas those using non-Latin scripts, such as CMN, THA, and JPN, tend to have lower translation quality. By comparing the translation quality in Table 14 with the recommendation performance in Table 16, we observe a general positive correlation between translation quality and the performance of translation-based baselines, namely DIRE-LLaMA (Trans. ML+Trans. ML) and DIRE-LLaMA (ENG+Trans. ML). This suggests that these baselines not only lack effective preference pattern transfer but are also heavily constrained by translation quality, resulting in inferior recommendation performance. In contrast, while PPT also utilizes translation during news augmentation, it effectively mitigates the semantic shift caused by poor translation quality through cross-lingual alignment and achieves strong performance across all languages.

| Language | BLEU | Language | BLEU |
|---|---|---|---|
| SWH | 0.3914 | THA | 0.1595 |
| SOM | 0.2469 | RON | 0.3267 |
| CMN | 0.1127 | FIN | 0.2121 |
| JPN | 0.1628 | KAT | 0.1919 |
| TUR | 0.2545 | HAT | 0.4199 |
| TAM | 0.2850 | IND | 0.3587 |
| VIE | 0.3108 | GRN | 0.2651 |

Table 14: Translation quality from minor language to English.

In our paper, the default translation model is the open-source NLLB-200-3.3B model. To further study the effects of translation quality, Table 15 compares the average performance of different translation models on xMIND. It can be observed that the choice of translation models has a limited impact on performance. Even when using the smallest model, NLLB-200-distilled-600M, PPT still outperforms all baselines, demonstrating its robustness to translation quality.

| Translation Model | BLEU | UAUC |
|---|---|---|
| NLLB-200-distilled-600M | 0.2020 | 59.13 |
| NLLB-200-distilled-1.3B | 0.2317 | 59.31 |
| NLLB-200-3.3B | 0.2336 | 59.39 |

Table 15: The effects of Translation Models. BLEU is calculated by back-translation.

| | SWH | | | | | THA | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** |
| DIRE-XLM-Roberta (ML+ML) | 56.67 | 29.29 | 14.47 | 29.79 | 36.02 | 56.86 | 29.05 | 13.56 | 29.57 | 36.00 |
| DIRE-LLaMA (ML+ML) | 56.10 | 30.09 | 15.28 | 30.90 | 37.02 | 53.56 | 28.68 | 12.71 | 29.79 | 35.68 |
| DIRE-LLaMA (Trans. ML+Trans. ML) | 57.80 | 31.22 | 15.65 | 32.43 | 38.37 | 57.64 | 30.25 | 15.64 | 31.99 | 38.06 |
| DIRE-XLM-Roberta (ENG+ML) | 57.51 | 31.66 | 17.63 | 32.28 | 38.25 | 55.76 | 29.26 | 14.53 | 29.88 | 36.06 |
| DIRE-LLaMA (ENG+ML) | 47.49 | 25.25 | 10.82 | 25.10 | 31.10 | 55.37 | 29.08 | 13.05 | 30.41 | 36.27 |
| DIRE-LLaMA (ENG+Trans. ML) | 58.26 | 31.03 | 15.40 | 32.26 | 38.38 | 57.66 | 30.58 | 15.46 | 31.73 | 37.85 |
| Cross-Lingual-NRMS | 58.01 | 31.67 | 16.71 | 32.64 | 39.04 | 54.93 | 29.04 | 13.66 | 30.04 | 36.37 |
| PPT | **60.82** | **34.02** | **19.44** | **35.29** | **41.04** | **57.88** | **31.17** | **15.85** | **32.49** | **38.42** |

| | SOM | | | | | RON | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** |
| DIRE-XLM-Roberta (ML+ML) | 57.29 | 29.80 | 14.86 | 30.35 | 36.79 | 57.59 | 29.53 | 13.77 | 30.44 | 36.61 |
| DIRE-LLaMA (ML+ML) | 55.25 | 28.57 | 14.31 | 29.43 | 35.21 | 57.13 | 29.75 | 15.24 | 30.23 | 36.62 |
| DIRE-LLaMA (Trans. ML+Trans. ML) | 57.66 | 31.13 | 15.35 | 32.44 | 38.29 | 58.41 | 31.26 | 16.49 | 32.55 | 38.12 |
| DIRE-XLM-Roberta (ENG+ML) | 54.60 | 28.56 | 13.69 | 29.32 | 35.56 | 55.49 | 30.41 | 15.60 | 29.99 | 36.87 |
| DIRE-LLaMA (ENG+ML) | 46.29 | 25.18 | 10.30 | 24.72 | 31.30 | 46.26 | 23.65 | 9.87 | 23.26 | 29.42 |
| DIRE-LLaMA (ENG+Trans. ML) | 56.49 | 30.10 | 15.22 | 30.88 | 37.14 | 56.76 | 30.77 | 15.75 | 31.31 | 37.60 |
| Cross-Lingual-NRMS | 55.69 | 30.27 | 14.85 | 31.38 | 37.77 | 58.50 | 31.63 | 15.53 | 33.12 | 39.20 |
| PPT | **58.46** | **32.50** | **17.35** | **33.91** | **39.83** | **61.39** | **33.98** | **18.07** | **35.77** | **41.30** |

| | CMN | | | | | FIN | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** |
| DIRE-XLM-Roberta (ML+ML) | 57.53 | 29.51 | 12.84 | 30.72 | 37.07 | 58.97 | 30.58 | 15.49 | 31.69 | 37.82 |
| DIRE-LLaMA (ML+ML) | 56.01 | 29.26 | 14.48 | 30.04 | 36.27 | 54.19 | 27.14 | 11.96 | 27.48 | 34.08 |
| DIRE-LLaMA (Trans. ML+Trans. ML) | 56.55 | 29.58 | 15.26 | 30.07 | 36.06 | 58.01 | 30.08 | 15.46 | 30.74 | 37.13 |
| DIRE-XLM-Roberta (ENG+ML) | 56.43 | 30.55 | 15.98 | 31.15 | 37.23 | 54.06 | 26.61 | 11.17 | 27.10 | 33.87 |
| DIRE-LLaMA (ENG+ML) | 54.58 | 29.59 | 13.84 | 30.68 | 36.55 | 45.77 | 25.14 | 11.11 | 24.42 | 30.58 |
| DIRE-LLaMA (ENG+Trans. ML) | 55.71 | 29.74 | 14.75 | 30.51 | 36.45 | 55.57 | 30.51 | 16.44 | 31.18 | 37.37 |
| Cross-Lingual-NRMS | 57.65 | 31.14 | 15.84 | 32.06 | 38.25 | 58.06 | 31.71 | 16.54 | 32.77 | 38.89 |
| PPT | **60.32** | **33.26** | **18.39** | **34.51** | **40.29** | **60.79** | **33.89** | **19.30** | **35.36** | **41.10** |

| | JPN | | | | | KAT | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** |
| DIRE-XLM-Roberta (ML+ML) | 56.63 | 29.87 | 15.36 | 30.12 | 36.57 | 54.11 | 26.95 | 12.41 | 27.27 | 33.68 |
| DIRE-LLaMA (ML+ML) | 56.79 | 29.47 | 15.09 | 30.46 | 36.60 | 53.26 | 28.40 | 12.84 | 29.66 | 35.54 |
| DIRE-LLaMA (Trans. ML+Trans. ML) | 56.39 | 29.74 | 14.91 | 30.52 | 36.65 | **55.60** | 29.15 | **14.92** | 29.01 | 35.92 |
| DIRE-XLM-Roberta (ENG+ML) | 56.00 | 29.85 | 14.62 | 30.70 | 36.69 | 52.09 | 26.46 | 12.28 | 26.14 | 33.11 |
| DIRE-LLaMA (ENG+ML) | 54.95 | 27.55 | 12.10 | 28.51 | 34.80 | 51.60 | 26.34 | 10.95 | 26.99 | 33.29 |
| DIRE-LLaMA (ENG+Trans. ML) | 58.04 | 31.69 | 16.81 | 32.97 | 38.71 | 55.11 | **29.79** | 14.57 | **30.86** | **36.89** |
| Cross-Lingual-NRMS | 56.45 | 30.37 | 15.65 | 31.43 | 37.55 | 51.96 | 27.17 | 11.56 | 27.51 | 34.38 |
| PPT | **59.41** | **32.70** | **18.16** | **33.88** | **39.85** | 54.59 | 29.03 | 13.47 | 29.89 | 36.19 |

| | TUR | | | | | HAT | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** |
| DIRE-XLM-Roberta (ML+ML) | 58.57 | 31.81 | 16.83 | 32.91 | 38.97 | 57.42 | 31.34 | 15.66 | 32.82 | 38.59 |
| DIRE-LLaMA (ML+ML) | 55.80 | 29.02 | 14.07 | 29.84 | 35.95 | 55.77 | 30.17 | 15.46 | 31.15 | 36.90 |
| DIRE-LLaMA (Trans. ML+Trans. ML) | 58.31 | 32.07 | 17.27 | 33.37 | 39.07 | 58.80 | 31.65 | 17.31 | 32.58 | 38.44 |
| DIRE-XLM-Roberta (ENG+ML) | 57.05 | 30.60 | 16.05 | 31.44 | 37.70 | 55.52 | 29.67 | 14.23 | 30.50 | 36.59 |
| DIRE-LLaMA (ENG+ML) | 48.40 | 25.67 | 12.15 | 25.06 | 31.57 | 45.78 | 23.96 | 8.74 | 24.10 | 30.09 |
| DIRE-LLaMA (ENG+Trans. ML) | 57.04 | 30.11 | 15.57 | 30.65 | 36.93 | 57.85 | 31.43 | 15.62 | 32.87 | 38.58 |
| Cross-Lingual-NRMS | 57.28 | 30.88 | 15.74 | 31.53 | 38.22 | 57.22 | 31.82 | 16.35 | 33.13 | 38.88 |
| PPT | **60.18** | **32.98** | **18.29** | **34.27** | **40.35** | **60.29** | **34.05** | **19.00** | **35.75** | **41.23** |

| | TAM | | | | | IND | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** |
| DIRE-XLM-Roberta (ML+ML) | 54.21 | 28.46 | 13.49 | 29.12 | 35.56 | 56.97 | 29.49 | 14.57 | 30.04 | 36.68 |
| DIRE-LLaMA (ML+ML) | 55.59 | 28.32 | 13.53 | 28.62 | 35.25 | 53.98 | 29.05 | 14.94 | 29.61 | 35.84 |
| DIRE-LLaMA (Trans. ML+Trans. ML) | 57.36 | 30.19 | 15.26 | 31.44 | 37.48 | 57.37 | 30.92 | 15.43 | 32.20 | 37.85 |
| DIRE-XLM-Roberta (ENG+ML) | 53.12 | 27.45 | 11.78 | 28.01 | 34.42 | 57.57 | 31.47 | 16.79 | 32.50 | 38.40 |
| DIRE-LLaMA (ENG+ML) | 53.48 | 27.93 | 11.85 | 28.99 | 34.79 | 45.85 | 25.01 | 10.80 | 24.69 | 30.61 |
| DIRE-LLaMA (ENG+Trans. ML) | 56.83 | 30.63 | 14.66 | 32.21 | 37.87 | 58.72 | 31.66 | 15.96 | 32.87 | 39.06 |
| Cross-Lingual-NRMS | 55.92 | 29.93 | 15.00 | 30.60 | 37.18 | 58.52 | 31.88 | 16.69 | 33.19 | 39.48 |
| PPT | **58.78** | **32.05** | **17.47** | **33.21** | **39.28** | **61.34** | **34.35** | **19.39** | **35.92** | **41.67** |

| | VIE | | | | | GRN | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** |
| DIRE-XLM-Roberta (ML+ML) | **58.53** | 31.14 | 16.03 | 32.16 | 38.49 | 55.89 | 29.00 | 13.29 | 30.14 | 36.37 |
| DIRE-LLaMA (ML+ML) | 55.01 | 30.72 | **16.77** | 31.56 | 37.23 | 56.58 | 30.19 | 15.52 | 30.78 | 36.90 |
| DIRE-LLaMA (Trans. ML+Trans. ML) | 57.33 | 31.14 | 16.25 | 32.28 | 38.02 | 57.01 | 30.51 | 15.71 | 31.45 | 37.27 |
| DIRE-XLM-Roberta (ENG+ML) | 56.66 | 30.74 | 15.59 | 31.73 | 37.68 | 53.98 | 30.00 | 15.92 | 30.20 | 36.97 |
| DIRE-LLaMA (ENG+ML) | 47.46 | 24.53 | 11.28 | 24.11 | 30.38 | 50.58 | 27.63 | 14.05 | 27.40 | 33.31 |
| DIRE-LLaMA (ENG+Trans. ML) | 57.53 | 31.54 | 16.22 | 32.79 | 38.61 | 55.52 | 29.68 | 14.79 | 30.28 | 36.50 |
| Cross-Lingual-NRMS | 54.82 | 29.84 | 14.26 | 31.20 | 37.15 | 56.88 | 30.38 | 15.29 | 31.09 | 37.65 |
| PPT | 57.84 | **32.03** | 16.50 | **33.71** | **39.29** | **59.33** | **32.59** | **17.80** | **33.58** | **39.65** |

Table 16: Performance comparison across all 14 minor languages. The best results are highlighted in **bold**. n@K denotes nDCG@K, similarly hereinafter.

| | SWH | | | | | THA | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** |
| / | **60.82** | **34.02** | **19.44** | **35.29** | **41.04** | **57.88** | **31.17** | **15.85** | **32.49** | **38.42** |
| Cross-Lingual News Augmentation | 60.21 | 33.57 | 19.25 | 34.86 | 40.73 | 57.49 | 30.87 | 15.37 | 32.24 | 38.06 |
| Cross-Lingual Alignment | 60.22 | 33.17 | 18.56 | 34.36 | 40.05 | 55.90 | 29.76 | 14.68 | 30.95 | 36.83 |
| Both Augmentation & Alignment | 58.17 | 31.09 | 16.21 | 31.85 | 38.15 | 55.83 | 29.73 | 14.60 | 30.74 | 36.72 |

| | SOM | | | | | RON | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** |
| / | 58.46 | 32.50 | 17.35 | 33.91 | 39.83 | **61.39** | 33.98 | 18.07 | 35.77 | 41.30 |
| Cross-Lingual News Augmentation | **58.92** | 32.38 | 17.24 | 33.88 | 39.49 | 61.27 | **34.49** | **19.30** | **35.90** | **41.59** |
| Cross-Lingual Alignment | 57.60 | **32.98** | **17.59** | **34.87** | **40.32** | 61.08 | 34.15 | 19.22 | 35.57 | 41.26 |
| Both Augmentation & Alignment | 56.73 | 30.48 | 15.09 | 31.54 | 37.76 | 61.01 | 33.35 | 18.06 | 34.81 | 40.75 |

| | CMN | | | | | FIN | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** |
| / | **60.32** | **33.26** | **18.39** | **34.51** | **40.29** | **60.79** | **33.89** | **19.30** | **35.36** | **41.10** |
| Cross-Lingual News Augmentation | 59.80 | 32.27 | 17.70 | 33.35 | 39.29 | 60.64 | 33.65 | 19.07 | 34.59 | 40.72 |
| Cross-Lingual Alignment | 57.08 | 31.07 | 16.48 | 32.39 | 37.92 | 60.52 | 33.81 | 19.00 | 35.30 | 40.99 |
| Both Augmentation & Alignment | 58.11 | 31.81 | 17.59 | 33.02 | 38.68 | 60.23 | 33.27 | 19.00 | 34.50 | 40.33 |

| | JPN | | | | | KAT | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** |
| / | 59.41 | **32.70** | **18.16** | 33.88 | **39.85** | **54.59** | 29.03 | 13.47 | **29.89** | **36.19** |
| Cross-Lingual News Augmentation | 59.08 | 32.31 | 17.73 | **33.91** | 39.60 | 53.59 | 28.74 | 13.46 | 29.22 | 35.20 |
| Cross-Lingual Alignment | 59.07 | 31.53 | 15.79 | 32.96 | 38.80 | 53.23 | **29.19** | **15.01** | 29.78 | 35.79 |
| Both Augmentation & Alignment | **59.60** | 31.87 | 16.89 | 33.00 | 38.94 | 53.82 | 28.80 | 14.47 | 29.22 | 35.11 |

| | TUR | | | | | HAT | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** |
| / | **60.18** | 32.98 | 18.29 | 34.27 | 40.35 | 60.29 | 34.05 | 19.00 | 35.75 | 41.23 |
| Cross-Lingual News Augmentation | 60.09 | 33.08 | 18.17 | **35.02** | **40.96** | **60.30** | **34.27** | **19.06** | **36.00** | **41.39** |
| Cross-Lingual Alignment | 60.12 | **33.18** | **18.41** | 34.35 | 40.20 | 59.43 | 33.08 | 18.09 | 34.55 | 39.97 |
| Both Augmentation & Alignment | 58.43 | 31.98 | 17.80 | 33.00 | 39.02 | 59.11 | 33.79 | 19.04 | 35.38 | 40.86 |

| | TAM | | | | | IND | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** |
| / | **58.78** | **32.05** | **17.47** | **33.21** | **39.28** | **61.34** | 34.35 | **19.39** | 35.92 | **41.67** |
| Cross-Lingual News Augmentation | 58.26 | 31.29 | 16.99 | 31.88 | 38.31 | 61.28 | **34.51** | 19.21 | **36.18** | 41.65 |
| Cross-Lingual Alignment | 57.57 | 31.23 | 16.58 | 32.15 | 38.24 | 59.66 | 31.14 | 16.05 | 32.21 | 38.53 |
| Both Augmentation & Alignment | 57.10 | 28.95 | 14.57 | 29.64 | 35.98 | 58.71 | 30.29 | 15.03 | 31.36 | 37.92 |

| | VIE | | | | | GRN | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** |
| / | 57.84 | 32.03 | 16.50 | 33.71 | 39.29 | **59.33** | **32.59** | 17.80 | 33.58 | **39.65** |
| Cross-Lingual News Augmentation | **58.50** | **32.67** | **17.04** | **34.60** | **40.04** | 59.19 | 32.38 | 17.84 | **33.59** | 39.62 |
| Cross-Lingual Alignment | 58.32 | 32.08 | 16.98 | 33.47 | 39.22 | 58.77 | 32.19 | 17.89 | 33.34 | 39.18 |
| Both Augmentation & Alignment | 56.57 | 31.28 | 16.48 | 32.58 | 38.18 | 58.69 | 32.28 | **18.15** | 33.35 | 39.24 |

Table 17: The ablation study across all 14 minor languages. The best results are highlighted in **bold**.

| | SWH | | | | | THA | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** |
| LLaMA-2-7b | **60.82** | **34.02** | **19.44** | **35.29** | **41.04** | **57.88** | **31.17** | **15.85** | **32.49** | **38.42** |
| BERT-base | 55.34 | 30.09 | 14.36 | 30.90 | 37.11 | 51.46 | 27.96 | 14.36 | 27.68 | 34.18 |
| XLM-RoBERTa-base | 56.40 | 30.34 | 15.06 | 31.50 | 37.52 | 56.90 | 30.68 | 15.60 | 31.51 | 37.80 |
| | **SOM** | | | | | **RON** | | | | |
| **Model** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** |
| LLaMA-2-7b | **58.46** | **32.50** | **17.35** | **33.91** | **39.83** | **61.39** | **33.98** | **18.07** | **35.77** | **41.30** |
| BERT-base | 54.05 | 28.73 | 12.65 | 29.89 | 35.85 | 55.58 | 30.07 | 14.16 | 31.30 | 37.15 |
| XLM-RoBERTa-base | 55.76 | 29.75 | 14.33 | 30.74 | 36.87 | 57.96 | 30.74 | 15.37 | 31.72 | 38.17 |
| | **CMN** | | | | | **FIN** | | | | |
| **Model** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** |
| LLaMA-2-7b | **60.32** | **33.26** | **18.39** | **34.51** | **40.29** | **60.79** | **33.89** | **19.30** | **35.36** | **41.10** |
| BERT-base | 54.32 | 29.01 | 13.43 | 29.72 | 35.96 | 56.11 | 31.62 | 16.54 | 33.06 | 38.67 |
| XLM-RoBERTa-base | 58.60 | 30.87 | 15.34 | 31.90 | 38.33 | 58.12 | 30.78 | 17.57 | 31.71 | 38.28 |
| | **JPN** | | | | | **KAT** | | | | |
| **Model** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** |
| LLaMA-2-7b | **59.41** | **32.70** | **18.16** | **33.88** | **39.85** | **54.59** | **29.03** | **13.47** | **29.89** | **36.19** |
| BERT-base | 54.37 | 27.77 | 12.39 | 28.13 | 34.91 | 51.86 | 27.50 | 11.70 | 27.63 | 34.39 |
| XLM-RoBERTa-base | 58.39 | 31.39 | 15.27 | 33.05 | 39.01 | 54.30 | 28.91 | 13.07 | 28.20 | 36.07 |
| | **TUR** | | | | | **HAT** | | | | |
| **Model** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** |
| LLaMA-2-7b | **60.18** | **32.98** | **18.29** | **34.27** | **40.35** | **60.29** | **34.05** | **19.00** | **35.75** | **41.23** |
| BERT-base | 55.74 | 29.11 | 13.19 | 29.99 | 36.38 | 58.15 | 31.64 | 15.44 | 32.98 | 38.94 |
| XLM-RoBERTa-base | 57.94 | 31.43 | 15.60 | 32.84 | 38.94 | 57.35 | 30.64 | 14.33 | 32.24 | 38.15 |
| | **TAM** | | | | | **IND** | | | | |
| **Model** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** |
| LLaMA-2-7b | **58.78** | **32.05** | **17.47** | **33.21** | **39.28** | **61.34** | **34.35** | **19.39** | **35.92** | **41.67** |
| BERT-base | 53.45 | 28.27 | 13.08 | 28.70 | 35.06 | 58.06 | 31.26 | 15.29 | 32.67 | 38.62 |
| XLM-RoBERTa-base | 56.70 | 30.35 | 14.71 | 31.48 | 37.65 | 57.90 | 29.74 | 14.34 | 31.11 | 37.43 |
| | **VIE** | | | | | **GRN** | | | | |
| **Model** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** | **UAUC** | **MRR** | **n@1** | **n@5** | **n@10** |
| LLaMA-2-7b | **57.84** | **32.03** | **16.50** | **33.71** | **39.29** | **59.33** | **32.59** | **17.80** | **33.58** | **39.65** |
| BERT-base | 55.71 | 29.68 | 13.83 | 30.98 | 36.87 | 56.82 | 31.13 | 16.01 | 31.83 | 37.87 |
| XLM-RoBERTa-base | 56.74 | 30.12 | 14.19 | 31.41 | 37.74 | 55.17 | 30.41 | 15.65 | 31.83 | 37.72 |

Table 18: Performance with different pre-trained model across all 14 minor languages on xMIND. The best results are highlighted in **bold**.