On the Effect of Incorporating Expressed Emotions in News Articles on Diversity within Recommendation Models

Mete Sertkan[1], Julia Neidhardt[1]

[1]Christian Doppler Laboratory for Recommender Systems, TU Wien, Vienna, Austria

Abstract

Despite news articles being highly edited and trimmed to maintain a neutral and objective tone, there are still stylistic residues of authors like expressed emotions, which impact the decision-making of users whether or not to consume the recommended articles. In this study, we delve into the effects of incorporating emotional signals within the *EmoRec* model on both emotional and topical diversity in news recommendations. Our findings show a nuanced alignment with users' preferences, leading to less diversity and potential creation of an "emotion chamber." However, it is crucial to model these emotional dimensions explicitly rather than implicitly as contemporary deep-learning models do. This approach offers the opportunity to communicate and raise awareness about the reduction in diversity, allowing for interventions if necessary. We further explore the complex distinction between intra-list and user-centric diversity, sparking a critical debate on guiding user choices. Overall, our work emphasizes the importance of a balanced, ethically-grounded approach, paving the way for more informed and diverse news consumption.

Keywords

recommender systems, news recommendation, emotion analysis, emotional diversity, topical diversity

1. Introduction

Personalized news recommenders are vital tools that help users navigate the overwhelming quantity of daily news, aiming to improve decision-making, conserve resources, and enhance satisfaction. These systems typically rely on content-based methods, considering not just semantic but also stylistic elements and emotions within news articles [1, 2]. People's decision-making is often influenced by emotional as well as rational factors [3], emphasizing the importance of recognizing and utilizing emotions in the recommendation process.

In [4], we focused on the expressed emotions within news content, proposing a multi-level emotion-aware news recommendation framework known as *EmoRec*. This model considers both the emotions contained within the titles and abstracts of news articles, and those aggregated across categories and subcategories. Through this approach, we found that incorporating emotions into recommendations led to performance gains, with certain nuances based on the

11th International Workshop on News Recommendation and Analytics (INRA), September 19th, 2023, co-located with the 17th ACM Conference on Recommender Systems, Singapore, Singapore.
mete.sertkan@tuwien.ac.at (M. Sertkan); julia.neidhardt@tuwien.ac.at (J. Neidhardt)
0000-0003-0984-5221 (M. Sertkan); 0000-0001-7184-1841 (J. Neidhardt)

granularity of emotion taxonomy and the level of information considered. However, we also noted that the inclusion of emotions might decrease the recommendations' emotional diversity.

Building on, in this paper, we extend our investigation into the impact of incorporating emotional signals on diversity, specifically examining both emotional diversity and topical diversity within the context of news recommendations. Different categories naturally possess varying emotional distributions, and it is logical to expect a decrease in emotional diversity with the alignment of recommendations when emotions are incorporated. Hence, we seek to explore to what extent this alignment occurs.

Moreover, we draw a crucial distinction between intra-list diversity (diversity within a recommendation list) and user-centric diversity (diversity in recommendations relative to a user's previous consumption behavior). This differentiation leads to a vital discussion on whether to provide a diverse recommendation list, leaving the choice to the user (intra-list), or to direct the user towards more diverse options (user-centric). By thoroughly understanding these facets, we aim to identify potential threats that might create a "emotion chamber." Ultimately, we strive to make users aware of these factors, empowering them to make more informed decisions and adopt more conscious consumption behavior.

2. Background

In this study, we delve into the impact of using emotional signals on the diversity of news recommendations by employing the previously introduced *EmoRec* model [4]. Recommender systems commonly utilize deep learning (DL) architectures, as they offer an end-to-end approach for extracting features, bypassing the need for manually crafted heuristics [5, 6]. This approach has proven particularly effective in the news recommendation field [7, 8, 9, 10], and *EmoRec* is aligned with this trend.

However, *EmoRec* stands out by explicitly modeling the emotional dimension, rather than implicitly incorporating all aspects of the given input, as common in typical DL models. This explicit consideration of emotions is vital, as it retains interpretability and recognizes that the stylistic properties of recommended items, including emotions, significantly influence user decision-making [3, 11].

Although emotions have been considered in recommender systems [11, 12, 13], their application in news recommendations is relatively unexplored. Emotions can be classified as expressed, perceived, or induced [11]. Our work, focusing on expressed emotions, adds a new dimension by extracting these emotions and incorporating them into the recommendation process.

Our work distinguishes itself by extensively examining three different emotion taxonomies and various levels of information, such as title, abstract, category, and subcategory. This goes beyond systems like [14, 15], which only explore sentiment, adding a novel and previously unexplored aspect to the news recommender system.

In summary, our work extends beyond *EmoRec*, an emotion-aware neural news recommendation model, to critically assess how the integration of emotions affects both emotional and topical diversity within recommended news articles. While *EmoRec* laid the foundation, our current study reaches beyond accuracy, opens new avenues, and provides valuable insights, especially in the complex field of news recommendation.

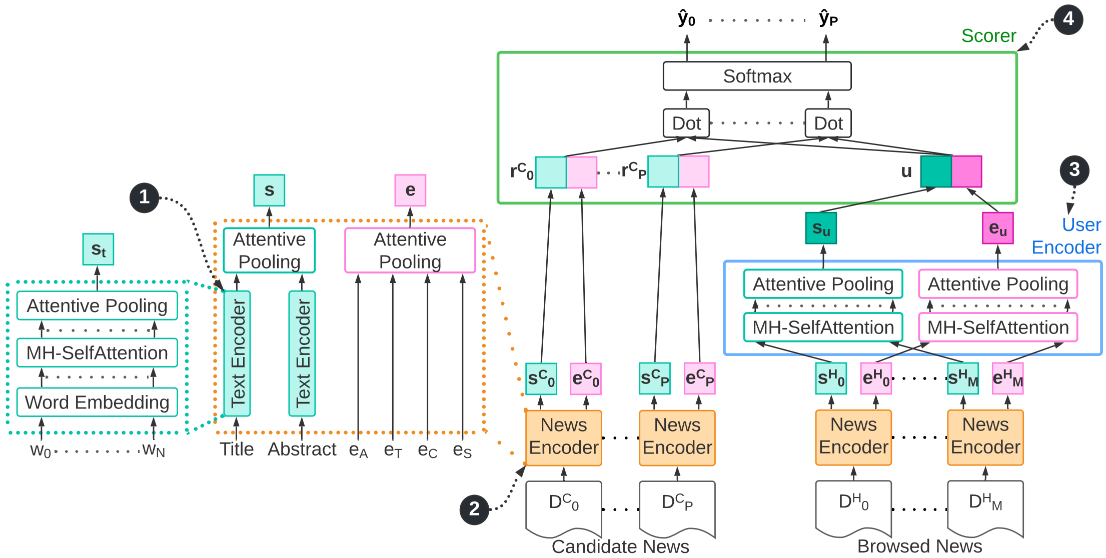

Figure 1: Our *EmoRec* Framework [4]: 1 *Text Encoder*, which learns a semantic representation s_t of any given sequence of words; 2 *News Encoder*, which utilizes the Text Encoder to obtain a semantic representation s of a news article by its title and abstract; and which combines the pre-computed emotional representations e_T, e_A, e_C, and e_S (i.e., title-, abstract-, category-, and subcategory-emotions) of a news article, to one representation e; 3 *User Encoder*, which separately models a semantic representation s_u and an emotional representation e_u of users based on their previous news interactions; 4 *Scorer*, which determines a score for a given user and candidate news pair. Note that the final representation of candidate news r_i^C and users u are simply the concatenation of their corresponding semantic and emotional representations.

3. Methods

3.1. Multi-Level Emotion-Aware News Recommender *EmoRec*

EmoRec [4] – illustrated in Figure 1 – is an emotion-aware news recommendation system. Its goal is to rank candidate news articles by considering both the user's interaction history and the emotional content of articles. *EmoRec* operates by analyzing a user's history of browsed news articles, and then it ranks a set of candidate articles by assigning a score to each one. Notably, the framework incorporates emotion scores from the articles in the recommendation process, considering emotions derived from the title, abstract, category, and subcategory of the articles. EmoRec also employs negative feedback to enhance its performance, learning from unclicked articles within a user's session. The model is trained to minimize the negative log-likelihood of the clicked news articles, with three different models being trained based on three distinct emotion taxonomies: Sentiment, Ekman, and GoEmotion (see Figure 2). For more details, readers are directed to our previous paper [4] and the corresponding repository[1] where *EmoRec* was first introduced.

[1] https://github.com/MeteSertkan/EmoRec

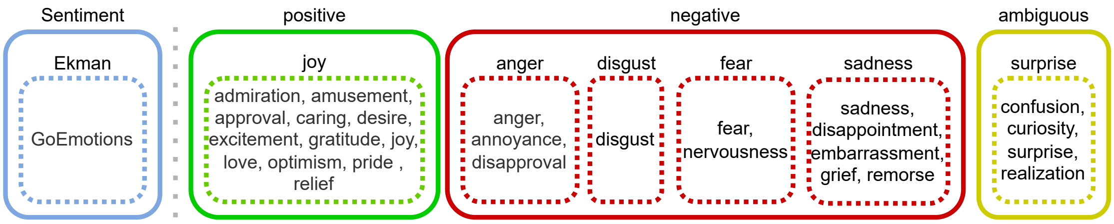

Figure 2: Emotion taxonomies [4] – On the left-hand side, we illustrate the hierarchical structure of the taxonomies and how to interpret the rest of the figure. Categories of the GoEmotions taxonomy [16] are mapped to the Ekman taxonomy [17] and to basic sentiments. For *example*, *anger*, *annoyance*, and *disapproval* map to *anger* and are overall *negative*. Note that neutral emotions are not listed here.

3.2. Diversity Metrics

We employ following diversity metrics: user-centric emotional diversity E_{UCD}, intra-list emotional diversity E_{ILD}, user-centric topical diversity T_{UCD}, and intra-list topical diversity T_{ILD}. While the intra-list diversity metrics compare news articles within the recommended list, the user-centric metrics put them in contrast to the users' previous consumption behavior. We take the cosine distance as the basis for our diversity metrics:

$$dist(v_{source}, v_{target}) = 1 - \frac{v_{source} \cdot v_{target}}{||v_{source}|| \ ||v_{target}||} \tag{1}$$

,

Depending on the computed metric, v_{source} and v_{target} are either emotion vectors of dimension 4, 7, or 28 (depending on the considered emotion taxonomy) or category embeddings of dimension 100.

Emotional Diversity. In comparing emotion-aware and non-emotion-aware recommendation models, we exclude the learned user emotion representation e_u and the weights used to combine various views such as title, abstract, category, and subcategory. We extract the emotion representation e^D_{TA} of a news article D using BERT-based classifiers, taking the title and abstract as input. This approach is consistent with the majority of baselines that rely solely on text. We then average emotion representations of all news articles in a user's history H to form the user's overall emotion representation, denoted as $\bar{e}_u$. Taking all this into account and given a ranked recommendation list L with R articles $[D_0, ..., D_R]$, we define the intra-list emotional diversity E_{ILD} as the average pairwise distance at cutoff K:

$$E_{ILD}@K = \frac{\sum_{D_i \in L@K} \sum_{D_j \in L@K \setminus \{D_i\}} dist(e^{D_i}_{TA}, e^{D_j}_{TA})}{K(K-1)}. \tag{2}$$

It provides insight into the emotional diversity of the ranked lists; greater diversity results in a higher $E_{ILD}@K$. Similarly, we define user-centric emotional diversity E_{UCD} as the average distance between the emotional representations of all news articles in the ranked recommendation list at cutoff K and the user's overall emotion orientation $\bar{e}_u$:

$$E_{UCD}@K = \frac{1}{K} \sum_{D_i \in L@K} dist(e_{TA}^{D_i}, \bar{e}_u). \tag{3}$$

It reflects how the ranked lists differ emotionally from the user's overall orientation. A greater difference in the top-K ranks results in higher values of $E_{UCD}@K$

Topical Diversity. We create 100 dimensional embeddings for categories c_C (e.g., for sports) and subcategories c_S (e.g., for soccer) of news articles. We average the (sub)category embeddings of all browsed news articles of users' to obtain their categorical representation c_u. Similarly, we average the (sub)category embeddings of top-K recommended news articles to obtain the recommendations category representation $c_{L@K}$. Having both, user-centric topical diversity is defined as:

$$T_{UCD}@K = dist(c_{L@K}, c_u), \tag{4}$$

indicating to what extent the consumed news articles differ from the recommended ones categorically (the higher the more diverse). For any given article D we compute its categorical representation c_D by averaging the article's category c_C and subcategory c_S embeddings. Therefore, we define the intra-list topical diversity T_{ILD} as the average pairwise distance at cutoff K:

$$T_{ILD}@K = \frac{\sum_{D_i \in L@K} \sum_{D_j \in L@K \setminus \{D_i\}} dist(c_{D_i}, c_{D_j})}{K(K-1)}, \tag{5}$$

which provides an intuition how the top-K ranked news articles diverge topically.

4. Experimental Setting

Our diversity analysis leverages the MIND-small[2] subset of the MIND dataset [10], specifically compiled from MSN News[3] logs collected between October 12 and November 22, 2019. The first five weeks of data were used for training, and the final week was allocated for testing. The dataset includes information from 50K randomly selected users who made at least five clicks, alongside 65K news articles, 230K impressions resulting in 350K clicks, and 8M instances where the users did not click. Each data sample consists of a timestamp, user ID, a chronologically arranged list of news IDs representing user interaction history, and a shuffled list of candidate news IDs labeled either as "clicked" or "seen but not clicked." In our study, we apply the *EmoRec* models trained in our earlier research; in particular *EmoRec*$_S$, *EmoRec*$_E$, *EmoRec*$_G$, where subscript refers to the used taxonomy (Sentiment, Ekman, or GoEmotion) for training. Specific details regarding the implementation, training, and fine-tuning of these models can be found in [4] and our *EmoRec*-repository[4]. We evaluate the diversity capabilities of *EmoRec* against several baseline models: LSTUR [1], a neural news recommendation system capturing users' long and

[2]https://msnews.github.io/index.html
[3]https://www.msn.com/en-us/news
[4]https://github.com/MeteSertkan/EmoRec

Table 1
Comparing emotional diversity (i.e., E_{UCD}@10 and E_{ILD}@10) of $EmoRec_S$, $EmoRec_E$, $EmoRec_G$, and the baselines. Subscripts S, E, and G indicate the used taxonomy for model training. Column names *Sentiment*, *Ekman*, and *GoEmotion* indicate the taxonomy used for distance calculation. Higher scores indicate more emotionally diverse recommendations. Note, † indicates a statistically significant difference to (our most emotionally diverse model) $EmoRec_E$ and ∗ indicates statistically significant difference to the random model, both at alpha 0.05.

	Model	Sentiment		Ekman		GoEmotion	
		E_{UCD}@10	E_{ILD}@10	E_{UCD}@10	E_{ILD}@10	E_{UCD}@10	E_{ILD}@10
1	Random	.1604†	.2378†	.1782†	.2665†	.2880†	.4348†
2	*SentiRec*	.1573∗†	.2341∗†	.1701∗†	.2560∗†	.2792∗†	.4214∗†
3	*NRMS*	.1607†	.2393†	.1729∗†	.2598∗†	.2883∗†	.4355∗†
4	*LSTUR*	.1666∗†	.2516∗†	.1762∗†	.2659†	.2859∗†	.4330∗†
5	*NAML*	**.1695**∗†	**.2559**∗†	**.1866**∗†	**.2818**∗†	**.3025**∗†	**.4611**∗†
6	$EmoRec_E$	.1616∗	.2424∗	.1720∗	.2591∗	.2822∗	.4280∗
7	$EmoRec_S$	.1584∗†	.2366∗†	.1686∗†	.2532∗†	.2772∗†	.4191∗†
8	$EmoRec_G$	.1570∗†	.2350∗†	.1675∗†	.2513∗†	.2742∗†	.4138∗†

short-term interests; NAML [6], which incorporates multiple views (title, abstract, category, and subcategory) into the news representation; NRMS [7], a neural news recommendation system employing multi-head self-attention for both news and user encoders; and SentiRec [8], a news recommender aware of sentiment diversity. Comprehensive details about these baseline models can be referred to in [14] and our *NewsRec*-repoistory[5]. We compare our results using paired *t*-tests with Bonferroni correction [18, 19].

5. Results

In this study, our primary focus is to investigate the impact of incorporating expressed emotions into the news recommendation process, particularly on diversity, Building up on our previous work [4] that introduced the *EmoRec* model. Previously [4], we uncovered that integrating emotions significantly enhances performance, with *EmoRec* surpassing all baseline models. Through a detailed analysis, we determined that both text-level emotions (derived from titles and abstracts) and category-level emotions (including those aggregated within subcategories) contributed to these improvements, with text-level emotions being the most influential. We also noted that using a broader emotion taxonomy yielded better results than a more nuanced one. In this work, while keeping the same settings and configuration, we shift our focus from merely improving accuracy to also understanding how these emotional elements effect diversity.

Emotional Diversity. In our analysis of emotional diversity among various models, we employ two specific evaluation metrics: E_{UCD} and E_{ILD} (details in Section 3). These diversity measures are influenced by the chosen emotion taxonomy (vector space), and therefore, we

[5]https://github.com/MeteSertkan/newsrec

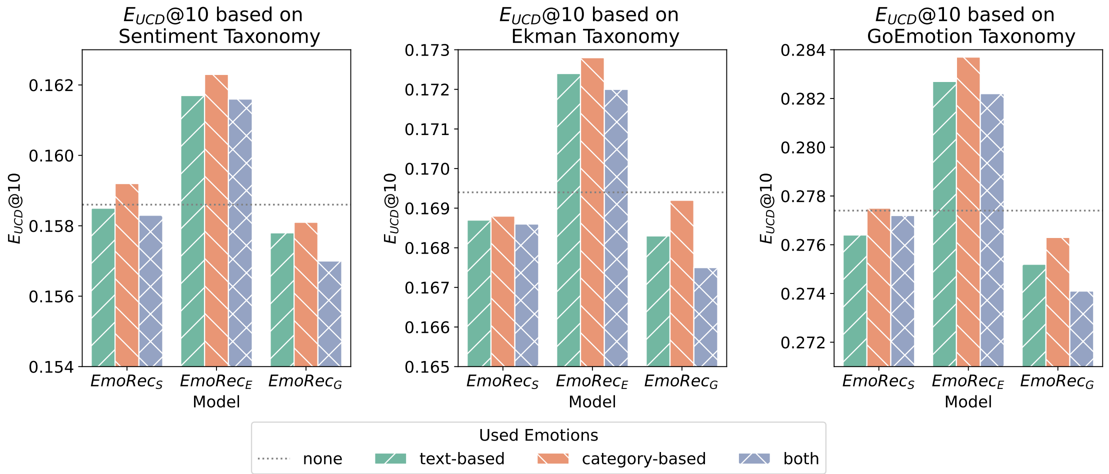

Figure 3: Ablation study on different configurations of *EmoRec* - User-centric emotional diversity analysis.

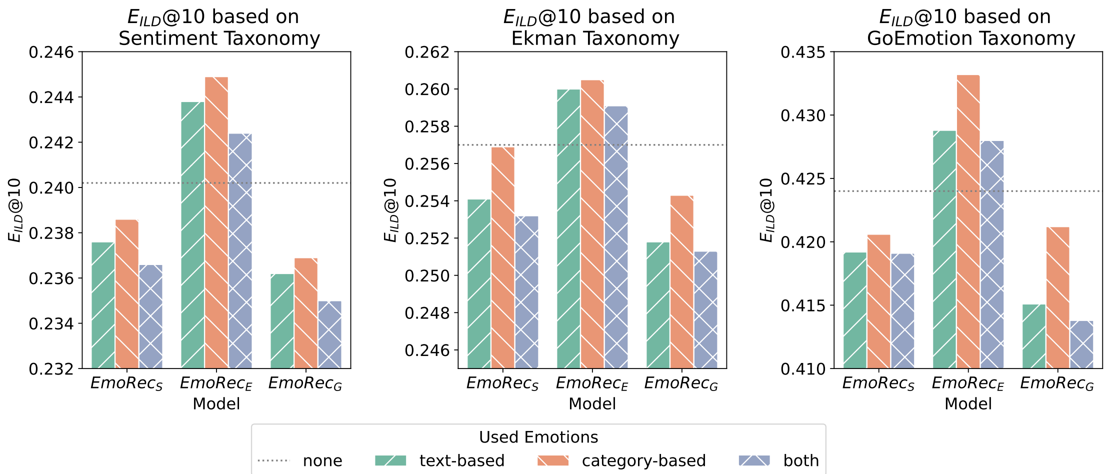

Figure 4: Ablation study on different configurations of *EmoRec* - Intra-list emotional diversity analysis.

calculate three distinct sets of emotional diversity metrics for each model. A summary of the results is presented in Table 1. Among the models, the *NAML* baseline demonstrates superior emotional diversity, surpassing all competitors, including the random model. Our models generally exhibit less emotional diversity in recommendations compared to purely text-based models (i.e., *NRMS*, *LSTUR*, *NAML*) with a few exceptions. The $EmoRec_E$ model only significantly outperforms *NRMS* and the random model in terms of emotional diversity when the *Sentiment* taxonomy is used.

In the ablation study, we assess our models' emotional diversity using four configurations: without emotions, utilizing text-based emotions, using category-based emotions, and incorporating both. Across user-centric and intra-list emotional diversity measures (see Figures 4

Table 2
Comparing topical diversity (i.e., T_{UCD}@10 and T_{ILD}@10) of $EmoRec_S$, $EmoRec_E$, $EmoRec_G$, and the baselines. Subscripts *S*, *E*, and *G* indicate the used taxonomy for model training. Higher scores indicate more topically diverse recommendations. Note, † indicates a statistically significant difference to $EmoRec_E$ and $*$ indicates statistically significant difference to the random model, both at alpha 0.05.

	Model	T_{UCD}@10	T_{ILD}@10
1	Random	**.5572†**	**.9225†**
2	*NRMS*	.5074∗†	.8984∗†
3	*SentiRec*	.5109∗†	.8986∗†
4	*LSTUR*	.5133∗†	.8974∗†
5	*NAML*	.5563∗†	.8367∗†
6	$EmoRec_E$	.4997∗	.8883∗
7	$EmoRec_G$	.4968∗	.8895∗
8	$EmoRec_S$	.4962∗	.8869∗

and 3), the models exhibit similar behavior. Our results show that the integration of emotions typically diminishes emotional diversity. However, an exception is found in $EmoRec_E$, where the inclusion of emotions enhances the diversity of recommendations. Furthermore, we consistently find that recommendations driven by category-based emotions are more diverse than those informed by text-based emotions. Interestingly, the model's full capacity configuration results in the least diverse recommendations.

6. Topical Diversity

In addition to emotional diversity, we also explore the topical alignment of the recommended items when emotions are incorporated. We evaluate topical diversity using both a user-centric metric (T_{UCD}) and an intra-list metric (T_{ILD}). The results are summarized in Table 2. The random model consistently provides the most diverse recommendations according to both metrics. In the context of user-centric emotional diversity, all *EmoRec* variants perform significantly worse than all baselines. We find a parallel trend in intra-list topical diversity, although in this instance, the *NAML* baseline performs even more poorly.

Figure 5 illustrates the topical diversity ablations – we use the same configurations as previously. The evaluation of user-centric topical diversity reveals a decrease in diversity across all configurations that incorporate emotions, whether text-based, category-based, or both. No specific pattern emerges to differentiate the effects of including text-based versus category-based emotions. When considering intra-list topical diversity, the inclusion of text-based emotions consistently leads to less diversity. However, in the models $EmoRec_E$ and $EmoRec_G$, this decrease is counterbalanced when category-based emotions are included. This results in a more topically diverse recommendation list in the full model, compared to configurations without emotions.

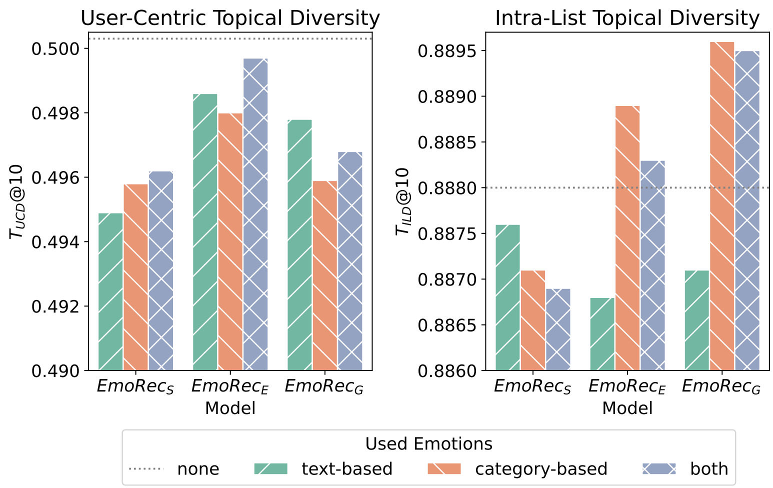

Figure 5: Ablation study on different configurations of *EmoRec* - User-centric & intra-list topical diversity analysis.

7. Discussion & Conclusions

News articles, often professionally edited to maintain a neutral tone, present unique challenges for recommendation systems. In our study, utilizing the MIND dataset [10], we observe that most articles lean towards a neutral score. However, our model, *EmoRec*, is designed to understand and exploit the subtle emotional variations within news articles, aligning them with users' consumption behavior to deliver more accurate recommendations.

We investigate the impact of incorporating emotional signals on diversity, both emotional and topical, within news recommendation models. Though *EmoRec* provides better alignment with users' preferences and yields higher accuracy, it also leads to a significant drop in diversity compared to other baselines. This reduction in diversity raises critical concerns about the potential creation of a self-reinforcing "emotion chamber" over time.

Deep-learning models, increasingly used in recommenders [5], implicitly account for textual nuances and users' tastes. The more proficient these models become, the more they align with users' preferences, potentially further reducing diversity. *EmoRec*, by explicitly modeling emotional dimensions, offers an opportunity to not only communicate and raise awareness about this issue but also intervene when necessary.

A critical aspect of our study involves distinguishing between intra-list diversity (within a recommendation list) and user-centric diversity (relative to a user's previous consumption behavior). This leads to a debate about the approach to recommendations: Should we provide users with diverse options and let them choose, or should we guide them towards more varied content? If the latter, what ethical considerations arise, such as justifying the recommendation of negative news following excessive positive consumption? We also contemplate a more nuanced approach, offering diverse options coupled with insights into a user's overall consumption behavior, enabling more informed decisions.

A recognized limitation in emotion-aware recommenders is the conflation of expressed, perceived, and induced emotions [11]. There are distinct differences between an article's emotional content, how users perceive that emotion, and the emotion actually induced in the reader. Moreover, the automated extraction process we employ adds a layer of complexity. We also urge caution in accepting established emotion taxonomies such as Ekman's [17], as they are highly debated and may be outdated [20]. These issues lead to the overarching question: What exactly are we measuring or considering with the extracted emotions?

In conclusion, our work highlights the complex interplay between accuracy and diversity in emotion-aware news recommendation. While *EmoRec* shows promising results, our findings emphasize the need for a thoughtful and ethically grounded approach to both user choice and emotional representation. In future work, we intend to investigate and compare different intervention strategies and delve into the nuanced differences in expressing, extracting, perceiving, and inducing emotions, as well as critically evaluate the taxonomies employed. This direction will help refine the alignment between user preferences and recommendations, facilitating more diverse and conscious consumption, without sacrificing the quality of the recommendations.

Acknowledgments

This research is supported by the Christian Doppler Research Association (CDG), and has received funding from the EU's H2020 research and innovation program (Grant No. 822670).

References

[1] M. Karimi, D. Jannach, M. Jugovac, News recommender systems – survey and roads ahead, Information Processing & Management 54 (2018) 1203–1227. URL: https://www.sciencedirect.com/science/article/pii/S030645731730153X. doi:https://doi.org/10.1016/j.ipm.2018.04.008.

[2] M. Sertkan, J. Neidhardt, H. Werthner, Documents, topics, and authors: Text mining of online news, in: 2019 IEEE 21st Conference on Business Informatics (CBI), volume 01, 2019, pp. 405–413. doi:10.1109/CBI.2019.00053.

[3] D. Jannach, M. Zanker, A. Felfernig, G. Friedrich, Online consumer decision making, Cambridge University Press, 2010, p. 234–252. doi:10.1017/CBO9780511763113.012.

[4] M. Sertkan, J. Neidhardt, Exploring expressed emotions for neural news recommendation, in: Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization, UMAP '22 Adjunct, Association for Computing Machinery, New York, NY, USA, 2022, p. 22–28. URL: https://doi.org/10.1145/3511047.3536414. doi:10.1145/3511047.3536414.

[5] S. Zhang, L. Yao, A. Sun, Y. Tay, Deep learning based recommender system: A survey and new perspectives, ACM Comput. Surv. 52 (2019). URL: https://doi.org/10.1145/3285029. doi:10.1145/3285029.

[6] Y. Deldjoo, M. Schedl, P. Cremonesi, G. Pasi, Recommender systems leveraging multimedia content, ACM Comput. Surv. 53 (2020). URL: https://doi.org/10.1145/3407190. doi:10.1145/3407190.

[7] M. An, F. Wu, C. Wu, K. Zhang, Z. Liu, X. Xie, Neural news recommendation with long- and short-term user representations, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Florence, Italy, 2019, pp. 336–345. URL: https://www.aclweb.org/anthology/P19-1033. doi:10.18653/v1/P19-1033.
[8] C. Wu, F. Wu, S. Ge, T. Qi, Y. Huang, X. Xie, 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), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 6389–6394. URL: https://www.aclweb.org/anthology/D19-1671. doi:10.18653/v1/D19-1671.
[9] C. Wu, F. Wu, M. An, J. Huang, Y. Huang, X. Xie, Neural news recommendation with attentive multi-view learning, arXiv preprint arXiv:1907.05576 (2019).
[10] F. Wu, Y. Qiao, J.-H. Chen, C. Wu, T. Qi, J. Lian, D. Liu, X. Xie, J. Gao, W. Wu, M. Zhou, MIND: A large-scale dataset for news recommendation, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 3597–3606. URL: https://www.aclweb.org/anthology/2020.acl-main.331. doi:10.18653/v1/2020.acl-main.331.
[11] E. Lex, D. Kowald, P. Seitlinger, T. N. T. Tran, A. Felfernig, M. Schedl, et al., Psychology-informed recommender systems, Now Publishers, 2021.
[12] Y. Zheng, B. Mobasher, R. D. Burke, The role of emotions in context-aware recommendation., Decisions@ RecSys 2013 (2013) 21–28.
[13] M. Tkalcic, A. Kosir, J. Tasic, Affective recommender systems: the role of emotions in recommender systems, in: Proc. The RecSys 2011 Workshop on Human Decision Making in Recommender Systems, Citeseer, 2011, pp. 9–13.
[14] M. Sertkan, S. Althammer, S. Hoftstätter, J. Neidhardt, Diversifying sentiments in news recommendation, in: Proceedings of the 2nd Perspectives on the Evaluation of Recommender Systems Workshop, 2022.
[15] C. Wu, F. Wu, T. Qi, Y. Huang, SentiRec: Sentiment diversity-aware neural news recommendation, in: Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing, Association for Computational Linguistics, Suzhou, China, 2020, pp. 44–53. URL: https://www.aclweb.org/anthology/2020.aacl-main.6.
[16] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, S. Ravi, GoEmotions: A Dataset of Fine-Grained Emotions, in: 58th Annual Meeting of the Association for Computational Linguistics (ACL), 2020.
[17] P. Ekman, Are there basic emotions? (1992).
[18] J. Urbano, H. Lima, A. Hanjalic, Statistical significance testing in information retrieval: An empirical analysis of type i, type ii and type iii errors, in: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'19, Association for Computing Machinery, New York, NY, USA, 2019, p. 505–514. URL: https://doi.org/10.1145/3331184.3331259. doi:10.1145/3331184.3331259.
[19] M. D. Smucker, J. Allan, B. Carterette, A comparison of statistical significance tests for information retrieval evaluation, in: Proceedings of the Sixteenth ACM Conference

on Conference on Information and Knowledge Management, CIKM '07, Association for Computing Machinery, New York, NY, USA, 2007, p. 623–632. URL: https://doi.org/10.1145/1321440.1321528. doi:10.1145/1321440.1321528.

[20] I. Nalis, J. Neidhardt, Not facial expression, nor fingerprint – acknowledging complexity and context in emotion research for human-centered personalization and adaptation, in: Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization, UMAP '23 Adjunct, Association for Computing Machinery, New York, NY, USA, 2023, p. 325–330. URL: https://doi.org/10.1145/3563359.3596990. doi:10.1145/3563359.3596990.