

Semantic Sampling: Enhancing Recommendation Diversity and User Engagement in the Headspace Meditation App

Rohan Singh Rajput*, Christabelle Pabalan*, Akhil Chaturvedi*, Prathamesh Kulkarni, Adam Brownell
Headspace
Los Angeles, USA

Abstract—In this paper, we present a clever approach to enhance the performance of sequential recommendation systems, specifically in the context of meditation recommendations within the Headspace app. Our method, termed “Semantic Sampling”, leverages the power of language embeddings and clustering techniques to introduce diversity and novelty in the recommendations. We augment the Time Interval Aware Self-Attention for Sequential Recommendation (TiSASRec) model with semantic sampling, where the next recommended item is randomly sampled from a cluster of semantically similar items. Our empirical evaluation, conducted on a sample set of 276,700 users, reveals a statistically significant increase of 2.26 % in content start rate for the treatment group (TiSASRec with semantic sampling) compared to the control group (TiSASRec alone). Furthermore, our approach demonstrates improved coverage and rarity, indicating a broader range of recommendations and higher novelty. The results underscore the potential of Semantic Sampling in enhancing user engagement and satisfaction in recommendation systems.

Keywords—Information retrieval; machine learning; recommender system.

I. INTRODUCTION

Recommendation systems have become an indispensable part of modern digital applications, providing users with personalized content and product suggestions [1]. These systems are especially important in the realm of wellness applications, where personalization can enhance user engagement and satisfaction significantly. Headspace, for instance, offers a rich variety of content covering meditation, sleep, focus, and music, among others [2].

However, a common challenge for these systems is the ‘long tail problem’. A significant number of items are rarely recommended, resulting in limited diversity in the recommendations [3]. This lack of diversity can curb user exploration and engagement, especially in areas where exploring new content is beneficial. In the context of wellness applications like Headspace, users gain exposure to a diverse range of content, aiding them in discovering new techniques, sustaining interest, and deepening their practice.

Unlike platforms like TikTok or YouTube, where content consumption can be sporadic and unplanned, Headspace users often embark on sequential mindfulness journeys through the content. Recognizing this unique behavior, we found TiSASRec, a sequential recommendation system built on the

transformer architecture, to be an apt baseline for our approach [4]. While TiSASRec outperformed earlier models, it started to loop back to repetitive content recommendations after extended training. This tendency to lean towards shorter, frequently accessed content over longer, seldom accessed content limited the scope of recommendations and potentially restricted user exploration.

To amplify recommendation diversity while retaining relevance, we integrated semantic sampling into the TiSASRec model. Semantic sampling involves extracting language embeddings from the titles and teasers of content using sentence transformers. After these embeddings are secured, we compute their cosine similarities. Instead of merely relying on TiSASRec’s recommendations, we select content from the cluster of the N most semantically similar items [5]. This approach not only broadens the array of recommendations but ensures they remain pertinent to the user, enhancing the discovery experience.

By melding sequential recommendations with semantic sampling, we present a robust solution to the long tail problem. This strategy augments the diversity of recommendations while ensuring their relevance to individual users. By blending the strengths of both sequential recommendation systems and semantic sampling, our goal is to elevate the user experience in wellness platforms like Headspace.

II. METHOD

To tackle the long-tail problem in recommendation systems, we introduced a technique called semantic sampling. This approach leverages language embeddings and cosine similarity to offer diverse and engaging content recommendations. We conducted extensive online A/B testing, evaluating various metrics to validate the efficacy of our approach. The results from the A/B test provided crucial insights into the real-world performance of our recommendation system.

A. Headspace App

Headspace is a popular mindfulness and meditation app that offers a wide variety of content to its users. The content is organized into different modules, each focusing on a specific topic or theme. The app provides recommendations to users on the “Today” tab, which is the main landing page of the app. The recommendations are personalized based on the user’s past interactions and preferences [6].

*Corresponding authors.

The app also provides recommendations in specific modules. For example, in the “Sleep” module, the app recommends sleep-focused content, while in the “Meditation” module, it recommends meditation-focused content. The goal of these recommendations is to provide users with content that is relevant and engaging, enhancing their overall experience with the app. This is further exemplified in Fig. 1.

B. Semantic Sampling

Semantic sampling is an approach we introduced to enhance the diversity of recommendations. The method employs the extraction of language embeddings from content titles and teasers using the Language-agnostic BERT Sentence Embedding (LaBSE) transformer model [7].

LaBSE is a multilingual sentence encoder, trained on a broad corpus of bilingual sentence pairs. It produces language-agnostic sentence embeddings, ensuring that sentences with equivalent meanings across different languages are proximal in the embedding space [7]. Such a property becomes invaluable for our use case, enabling the comparison and identification of similarities between content, irrespective of language barriers.

Our choice of LaBSE was driven by its excellent performance in paraphrasing similarity tasks, vital for our objective of discerning semantically similar content pieces. The illustration of Semantic Sampling workflow is explained in the Fig. 2.

Semantic sampling proceeds through the following stages:

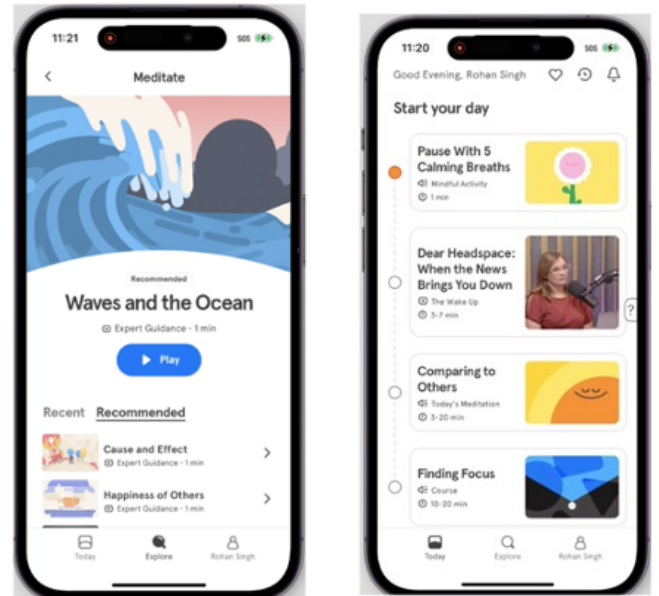
Content Embedding: We commence by extracting the language embeddings from content titles and teasers using LaBSE. This yields a high-dimensional vector representation for every content item.

Similarity Calculation: Cosine similarity between these embeddings is computed. This metric quantifies the cosine of the angle between two vectors, representing a measure of their alignment and, by extension, their similarity. Its value spectrum ranges from -1 (entirely dissimilar) to 1 (perfectly similar).

Recommendation Refinement: Instead of directly using the recommendation produced by the TiSASRec model, we assess the similarity of this chosen content with all other content items. Content pieces falling below a cosine similarity threshold of 0.75 to the chosen content are filtered out. From the remaining, more similar content pieces, one is selected randomly. For this process, there’s an 80% likelihood that the content will be resampled using semantic sampling.

The TiSASRec model underpins our recommendation system. A transformer-centric model, it has been empirically validated to adeptly capture users’ sequential behaviors [8]. The model utilizes the transformer framework, rooted in self-attention mechanisms, to emulate the sequential tendencies of users. This endows it with the capability to understand both immediate and extended user preferences, rendering it apt for our application.

For our endeavors, the TiSASRec model was the source of initial recommendations, which subsequently received enhancement through our semantic sampling technique.



(a) Recommended Meditation

(b) Today Tab

Fig. 1. Screenshots of the Headspace app showcasing our recommendation system. (a) The ‘Recommended for You’ section in the Meditation module displays personalized suggestions. (b) The ‘Today Tab’ features dynamic content shelves, each filled with diverse recommendations from our system for an engaging experience.

$$\text{Semantic Sampling}(c) = \arg \max_{c' \in C} \frac{1}{N} \sum_{i=1}^N \text{sim}(c, c'_i) \quad (1)$$

Here, c symbolizes the content piece put forth by the TiSASRec model, C embodies the collection of all content items, N stands for the count of the most similar items taken into account, and $\text{sim}(c, c'_i)$ signifies the cosine similarity between the embeddings of content pieces c and c'_i .

By leveraging this methodology, we facilitate a richer spectrum of recommendations. This not only bolsters the user’s exploratory experience but also ensures the continued relevance of recommendations to individual users.

C. Metrics for Evaluation

In the context of recommender systems, research has traditionally focused on the precision of the recommendations. However, it has been recognized that other recommendation qualities—such as whether the list of recommendations is diverse and whether it contains novel items—may have a significant impact on the overall quality of a recommender system. Consequently, the focus of recommender systems research has shifted to include a wider range of ‘beyond accuracy’ objectives [9]. These metrics include coverage, entropy, rarity, and intra-list diversity (ILD).

1) Coverage: Coverage is a measure of the proportion of items in the catalog that the recommender system can suggest. It provides an understanding of how well the recommendations cover the available items. A higher coverage indicates that the recommender system is capable of suggesting a wider

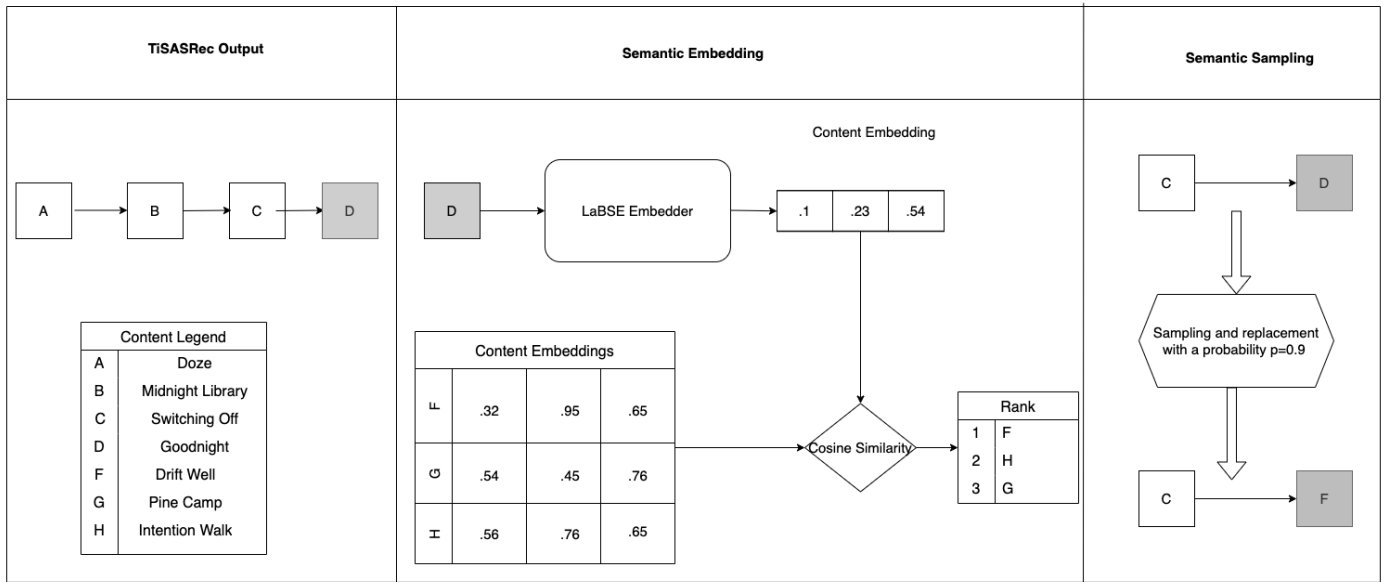


Fig. 2. Semantic Sampling Workflow. (a) The left side illustrates the input-output sequence of the TiSASRec model. (b) The central block depicts the generation of the LaBSE Embedding. (c) The right block demonstrates the replacement of the TiSASRec output with content that has been semantically embedded and sampled.

variety of items, which can contribute to a more diverse and personalized user experience [10]. It can be calculated as follows:

$$Coverage = \frac{|I_{rec}|}{|I|} \quad (2)$$

where, I_{rec} is the set of items recommended and I is the total set of items.

2) *Entropy*: Entropy is a measure of the unpredictability or randomness of the recommendations. It is derived from information theory, where it is used to quantify the amount of information contained in a set of data. In the context of recommender systems, a higher entropy indicates a more diverse set of recommendations, as it suggests that the recommendations are spread out over a larger number of different items. Optimal entropy is achieved when the recommendation distribution is uniformly distributed, therefore, an increase in entropy signifies an improvement for the long-tail problem [11]. It can be calculated as follows:

$$Entropy = - \sum_{i=1}^{|I|} p(i) \log p(i) \quad (3)$$

where, $p(i)$ is the probability of item i being recommended.

3) *Rarity*: Rarity is a measure of how uncommon or unique the recommended items are. It is defined as the inverse of normalized popularity, with 0 being our most viewed content and 1 being our least viewed content. A higher rarity score indicates that the recommender system is suggesting more unique or less popular items, which can contribute to a more diverse set of recommendations [12]. It can be calculated as follows:

$$Rarity = 1 - \frac{pop(i)}{\max_{j \in I} pop(j)} \quad (4)$$

where, $pop(i)$ is the popularity of item i , and $\max_{j \in I} pop(j)$ is the popularity of the most popular item.

4) *Intra-List Diversity (ILD)*: Intra-List Diversity (ILD) is a metric that measures the average dissimilarity between all pairs of items within a recommendation list. It is a measure of the diversity of the recommendations and is defined as follows:

$$ILD(L) = \frac{2}{|L| \cdot (|L| - 1)} \sum_{i=1}^{|L|} \sum_{j=i+1}^{|L|} d(i, j) \quad (5)$$

where, L is the list of recommended items, $|L|$ is the number of items in the list, and $d(i, j)$ is the dissimilarity between items i and j . The dissimilarity between items can be calculated using various methods, such as cosine distance in the embedding space. A higher ILD value indicates a more diverse set of recommendations [13].

5) *Content Click-Through Rate*: Content Click-Through Rate (CTR) is a measure of user engagement with the recommended content. It is defined as the number of times users initiate interaction with the content divided by the number of times the content is displayed to the users. This metric directly reflects the user's interaction with the recommended content, providing a clear measure of the effectiveness of the recommendations. An increase in this rate is indicative of users finding the recommendations more engaging and relevant [14]. It can be calculated as follows:

$$CTR = \frac{Clicks}{Impressions} \quad (6)$$

where *Clicks* is the number of times users initiate interaction with the content, and *Impressions* is the number of times the content is displayed to the users.

D. Online Analysis

For our online analysis, we conducted an A/B test on 138.4K control users (TiSASRec only) and 138.3K treatment users using the StatSig platform. StatSig is a platform that provides robust statistical analysis for A/B testing, ensuring that our results are statistically significant and reliable. This A/B test was run for a period of 42 days.

A key metric we focused on was the change in average content starts across the entire Headspace App. The content start rate, a widely accepted measure of user engagement, is defined as the number of times users initiate interaction with the content divided by the number of times the content is displayed to the users. This metric was chosen because it directly reflects the user's interaction with the recommended content, providing a clear measure of the effectiveness of the recommendations. An increase in this rate is indicative of users finding the recommendations more engaging and relevant [14].

To gain a deeper understanding of the changes in diversity brought about by our semantic sampling approach, we also calculated the entropy, ILD, coverage, and rarity for the clicks originating from both TiSASRec and the treatment model. These metrics were computed for the entire app, providing a comprehensive evaluation of the impact of semantic sampling on the quality of recommendations as described above.

This comprehensive online analysis allowed us to assess the real-world performance of our semantic sampling approach, providing valuable insights into its effectiveness in enhancing the diversity of recommendations while maintaining user engagement which is our primary goal.

III. RESULTS

In this section, we present the results of our study comparing the performance of the baseline TiSASRec model with the enhanced approach using semantic sampling. We evaluated the two models across various metrics to assess the impact on recommendation diversity and user engagement. Additionally, we discuss the findings from the online A/B test, which provided insights into the real-world effectiveness of our semantic sampling approach to solve the long tail problem.

A. Semantic Sampling Improves Recommendation Diversity

We conducted extensive evaluations to compare the performance of TiSASRec with our semantic sampling approach based on results from the online experiment. The results of these comparisons across different metrics are exemplified in Table I and Fig. 3.

1) *Coverage*: We first examined the coverage metric, which measures the proportion of items in the catalog that the recommender system is able to suggest. The results showed that the semantic sampling approach significantly outperformed TiSASRec across all top-k rankings (Coverage@1, Coverage@5, and Coverage@16). For example, at Coverage@1, semantic sampling achieved a coverage of 0.936, representing a substantial increase of 26.67% compared to TiSASRec's coverage

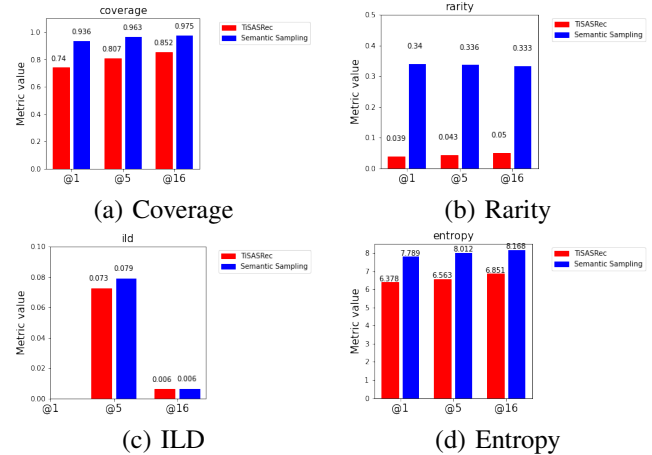


Fig. 3. Four figures depicting various recommender system metrics. (a) Coverage illustrates the algorithm's range of potential recommendations. (b) Rarity indicates the uniqueness of recommendations. (c) ILD represents the average dissimilarity between recommended items. (d) Entropy quantifies the information in a stochastic process.

of 0.740. This indicates that the semantic sampling approach recommended a wider variety of items to users, enhancing their opportunity for content discovery and engagement.

2) *Entropy*: The entropy metric measures the unpredictability or randomness of the recommendations. Higher entropy values suggest a more diverse set of recommendations. Our results revealed that the semantic sampling approach significantly increased the entropy of recommendations compared to TiSASRec. For instance, at Entropy@1, the mean entropy of semantic sampling was 7.789, which was 22.09% higher than TiSASRec's mean entropy of 6.378. Similarly, at Entropy@5 and Entropy@16, semantic sampling demonstrated improvements of 22.09% and 19.23%, respectively. These results indicate that the semantic sampling approach generated more diverse and less predictable recommendations, fostering a richer and more engaging user experience.

3) *Rarity*: The rarity metric measures how uncommon or unique the recommended items are. A higher rarity score indicates that the recommender system suggests more unique or less popular items, contributing to a more diverse set of recommendations. The semantic sampling approach significantly increased the rarity scores compared to TiSASRec for all top-k rankings (Rarity@1, Rarity@5, and Rarity@16). For example, at Rarity@1, semantic sampling achieved a score of 0.340, representing a remarkable increase of 765.88% compared to TiSASRec's score of 0.039. These results further confirm that semantic sampling effectively promotes the discovery of less frequently recommended content items.

4) *Intra-List Diversity (ILD)*: Intra-List Diversity (ILD) quantifies the average dissimilarity between all pairs of items within a recommendation list. Our semantic sampling approach achieved a modest improvement in ILD at top-k ranking values of ILD@5 and ILD@16, with a 8.82% and 2.84% increase, respectively. These results indicate a positive impact on the diversity of recommendations.

TABLE I. COMPARISONS BETWEEN TiSASREC AND SEMANTIC SAMPLING ACROSS DIFFERENT METRICS

Metric	TiSASRec		Semantic Sampling		% Change (Mean)
	Mean	Std	Mean	Std	Mean
Coverage@1	0.740	0.015	0.936	0.008	+26.670%
Coverage@5	0.807	0.010	0.963	0.004	+18.520%
Coverage@16	0.852	0.008	0.975	0.002	+13.550%
Entropy@1	6.378	0.039	7.789	0.026	+22.090%
Entropy@5	6.563	0.026	8.012	0.019	+22.090%
Entropy@16	6.851	0.019	8.168	0.013	+19.230%
Rarity@1	0.039	0.001	0.340	0.007	+765.880%
Rarity@5	0.043	0.001	0.336	0.006	+688.120%
Rarity@16	0.050	0.001	0.333	0.006	+562.830%
ILD@1	NaN	NaN	NaN	NaN	NaN
ILD@5	0.073	0.000	0.079	0.000	+8.820%
ILD@16	0.006	0.000	0.006	0.000	+2.840%

TABLE II. A/B TESTING EXPERIMENT RESULTS

Metric	Control	Treatment
Members	138.4K	138.3K
Average Content Starts	15.57	15.92
Delta % (Content Starts)	+2.26%	
p-value	0.004	

B. Online A/B Test Results

To validate the real-world impact of our semantic sampling approach, we conducted an A/B test involving a substantial user base of 276,700 members. The test compared the control group, which used the baseline TiSASRec model, with the treatment group, which experienced the enhanced model with semantic sampling. These results are shown in Table II.

1) *Average Content Starts*: As a media-oriented app, one of the primary metrics we focused on in the A/B test was the average content starts per user during the experimental period. This metric measures how frequently users initiated interactions with the recommended content. The treatment group, which used the semantic sampling approach, exhibited an average content start rate of 15.92 starts per user, while the control group, using the baseline TiSASRec model, had an average content start rate of 15.57 starts per user.

The 2.26% lift in content starts for the treatment group compared to the control group was statistically significant (p-value ≤ 0.05), indicating that the increase in user engagement with the recommended content was not due to random chance. This result highlights the effectiveness of the semantic sampling approach in encouraging users to interact more frequently with the content suggested by the recommender system.

IV. DISCUSSION

The results of our study demonstrate the effectiveness of the semantic sampling approach in enhancing the diversity of recommendations and increasing user engagement in the Headspace app. The evaluation of diversity metrics from the online experiment showed significant improvements in coverage, entropy, rarity, and intra-list diversity, indicating that the semantic sampling approach successfully addressed the long tail problem in recommendation systems. By suggesting more diverse, unique, and less predictable content to users, the semantic sampling approach enriches users' discovery

experience, encouraging them to explore a wider range of content.

The online A/B test further validated the real-world impact of the semantic sampling approach, showing a statistically significant lift of 2.26% in average content starts for the treatment group. This indicates that users in the treatment group found the recommendations generated using semantic sampling to be more engaging and relevant, leading to increased interactions with the recommended content.

A. Explanation of Increased Diversity and Relevance

The success of the semantic sampling approach in enhancing diversity and relevance can be attributed to two key factors:

Semantic Understanding: The use of language embeddings allowed the system to better understand the semantic meaning of content items. By capturing the inherent relationships between content titles and teasers, the approach could identify and group together semantically similar pieces. This understanding enabled the recommendation system to present a broader range of content options to users, encompassing items that share similar themes or topics.

Random Sampling: The introduction of random sampling from the cluster of semantically similar items injected diversity into the recommendation process. Instead of being confined to a fixed set of items, users were presented with randomly selected content pieces with similar meanings. This randomness allowed for serendipitous discoveries and introduced novelty, making the user experience more exciting and diverse.

Enhanced Relevance: Despite the introduction of diversity through random sampling, the semantic sampling approach ensured that the recommended items remained highly relevant to each individual user. By selecting items from the cluster of semantically similar content, the approach ensured that the recommendations retained a certain level of thematic coherence and alignment with users' preferences. This balance between diversity and relevance led to a more personalized and engaging experience for users, as they received a mix of both familiar and novel content that resonated with their interests.

Overall, the semantic sampling approach struck a delicate balance between increasing diversity and maintaining relevance, making it a powerful tool in addressing the long-tail

problem in recommendation systems. By leveraging semantic understanding and random sampling, the approach provided users with a diverse and personalized set of recommendations that enriched their discovery experience while ensuring the content remained highly relevant to their individual tastes and preferences.

B. Limitations

While the semantic sampling approach offers a promising solution to the long-tail problem and has demonstrated significant improvements in diversity and user engagement, it does have certain limitations that should be considered:

Limited Diversity Boost for Extremely Niche Content: The semantic sampling approach relies on identifying semantically similar content items to enhance diversity. However, for extremely niche or specialized content items that have limited semantic similarities with other items in the system, the approach may have limitations in boosting their diversity. This could lead to less diverse recommendations for such niche content.

Language Embedding Quality: The effectiveness of the semantic sampling approach is highly dependent on the quality of language embeddings obtained from models like LaBSE. Any limitations or biases present in the language embedding model can impact the accuracy of semantic similarities and, consequently, the diversity of recommendations. Ensuring the high quality and representativeness of language embeddings is critical for the success of the approach.

Impact of Sampling Parameters: The semantic sampling approach involves selecting a certain number of most semantically similar items (N) from which to sample recommendations. The choice of N can influence the level of diversity and relevance of the recommendations. Suboptimal values of N may lead to underemphasizing or overemphasizing certain content clusters, affecting the overall quality of recommendations. Careful experimentation and tuning of the sampling parameters are necessary for optimal results.

The semantic sampling approach represents a notable advancement in recommendation systems, with the potential to enhance diversity and user engagement in the wellness content domain. Ongoing research and refinement can address its limitations and further amplify its impact.

V. CONCLUSION

In this study, we introduced a semantic sampling method tailored for wellness recommendation systems, with an emphasis on the Headspace app. Online A/B testing involving over 276,700 users revealed that our method yielded a 2.26% uptick in the average content start rate—a clear indication of elevated user engagement.

In terms of diversity metrics, our method surpassed the TiSASRec baseline consistently. There were marked gains in metrics such as coverage, entropy, and rarity. Notably, the semantic sampling method ensured a broader range of engaging content recommendations without sacrificing user-specific relevance. Furthermore, increased content diversity did not detract from the start rate, underscoring the method's

capability to harmonize between tailored recommendations and content variety.

From a practical standpoint, semantic sampling amplifies the value proposition of content-rich platforms like Headspace. It fosters an environment where users are more inclined to explore diverse content, leading to a dynamic and individualized user journey. Future investigations might delve into refining similarity clustering methodologies and probing the long-term user satisfaction for lasting effects.

In summation, our research underscores semantic sampling's potential in augmenting both diversity and engagement in wellness recommendation systems, solidifying user satisfaction, and ensuring sustained app interaction.

ACKNOWLEDGMENT

This work was funded through Headspace's internal research and development budget. We thank Rachel Stevenson and Setu Shah all of whom are current employees of Headspace Health, for their feedback and contributions to this paper.

DATA AVAILABILITY

For this study, we used a dataset of Headspace App's user analytics events, specifically pertinent to content starts, plays, and completes on recommended content on all tabs and subsections of the app. This data is not publicly available due to privacy reasons and to safeguard Headspace's in-house analytics data.

REFERENCES

- [1] L. Lü, M. Medo, C. H. Yeung, Y.-C. Zhang, Z.-K. Zhang, and T. Zhou, "Recommender systems," *Physics reports*, vol. 519, no. 1, pp. 1–49, 2012.
- [2] M. Economides, J. Martman, M. J. Bell, and B. Sanderson, "Improvements in stress, affect, and irritability following brief use of a mindfulness-based smartphone app: a randomized controlled trial," *Mindfulness*, vol. 9, no. 5, pp. 1584–1593, 2018.
- [3] M. Kunaver and T. Požrl, "Diversity in recommender systems—a survey," *Knowledge-based systems*, vol. 123, pp. 154–162, 2017.
- [4] J. Li, Y. Wang, and J. McAuley, "Time interval aware self-attention for sequential recommendation," in *Proceedings of the 13th international conference on web search and data mining*, 2020, pp. 322–330.
- [5] Z. Xu, C. Chen, T. Lukaszewicz, Y. Miao, and X. Meng, "Tag-aware personalized recommendation using a deep-semantic similarity model with negative sampling," in *Proceedings of the 25th ACM International Conference on Information and Knowledge Management*, 2016, pp. 1921–1924.
- [6] R. L. Lee, "Review of headspace: Meditation and sleep." 2023.
- [7] F. Feng, Y. Yang, D. Cer, N. Arivazhagan, and W. Wang, "Language-agnostic bert sentence embedding," *arXiv preprint arXiv:2007.01852*, 2020.
- [8] J. Li, Y. Wang, and J. McAuley, "Time interval aware self-attention for sequential recommendation," *Proceedings of the Thirteenth ACM International Conference on Web Search and Data Mining*, p. 9, 2020.
- [9] P. Castells, N. Hurley, and S. Vargas, "Novelty and diversity in recommender systems," in *Recommender systems handbook*. Springer, 2021, pp. 603–646.
- [10] T. Zhou, Z. Kuscsik, J.-G. Liu, M. Medo, J. R. Wakeling, and Y.-C. Zhang, "Solving the apparent diversity-accuracy dilemma of recommender systems," *Proceedings of the National Academy of Sciences*, vol. 107, no. 10, pp. 4511–4515, 2010.
- [11] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Advances in artificial intelligence*, vol. 2009, 2009.

- [12] Y. C. Zhang, D. Ó. Séaghdha, D. Quercia, and T. Jambor, "Auralist: introducing serendipity into music recommendation," in *Proceedings of the fifth ACM international conference on Web search and data mining*, 2012, pp. 13–22.
- [13] C.-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen, "Improving recommendation lists through topic diversification," in *Proceedings of the 14th international conference on World Wide Web*, 2005, pp. 22–32.
- [14] J. Liu, P. Dolan, and E. R. Pedersen, "Personalized news recommendation based on click behavior," in *Proceedings of the 15th international conference on Intelligent user interfaces*, 2010, pp. 31–40.