4 EVALUATION OF INFORMATION COCOON

4.1 Experimental Setup

4.1.1 Dataset. We use the large-scale news recommendation benchmark dataset MIND [35] to conduct our experiments, which contains 6 weeks of news click records of anonymous users extracted from Microsoft News. In the data we used, there are a total of 2,609,219 behavior log records from 750,434 anonymous users. The news data provided in MIND contains rich text information, including: news ID, topic category, topic subcategory, title, abstract, URL, and entities. User information includes user ID and historical clicked news list. Each behavior log records the user's click behavior (click or not) on a group of news displayed on the platform at a certain moment. In addition, since the user behavior data provided in MIND's test set does not contain the label of whether the user clicked, we split half of the data from the validation set as the test set, and only use the remaining validation data to participate in model training. The statistics of the dataset are shown in Table 1.

Table 1: Statistics of MIND dataset.

MIND	train set	validation set	test set
# users	711,222	151,244	151,021
# news	101,527	72,023	72,023
# behavior logs	2,232,748	188,235	188,236

4.1.2 Algorithms. We explore the impact of three types of recommendation algorithms on the formation of information cocoons, including collaborative filtering-based recommendation algorithm, content-based recommendation algorithm, and hybrid recommendation algorithm. Under each type of recommendation algorithms, we selected some representative recommendation models, as follows:

Collaborative filtering-based recommendation algorithms. This type of algorithms make recommendations based on the click relationship between users and items. **NCF** [13]: Neural collaborative filtering recommendation model. NCF use a multi-layer perceptron to learn the interaction between users and items. **NGCF** [32]: Graph neural network based collaborative filtering recommendation model. NGCF constructs a graph structure of users and items, and learns embedding representations of users and items through graph neural networks.

Content-based recommendation algorithms. This type of algorithms make recommendations based on users' historical click data and the content of items. **NRMS** [34]: News recommendation model based on neural networks and self-attention mechanism. NRMS learns news representations from news titles and user representations from user clicked news list. **NAML** [33]: News recommendation model based on neural networks and multi-view learning. NAML learns news representations from three views: news titles, news content, and topic categories, and learns user representations from user clicked news list. **DKN** [30]: News recommendation model based on neural networks and knowledge graph representation. DKN uses knowledge-aware convolutional neural networks to fuse semantic-level and knowledge-level representations of news.

Hybrid recommendation algorithms. This type of algorithms combine multiple recommendation algorithms and can incorporate rich features for recommendation. **DeepFM** [11]: Factorization machine and neural networks based recommendation model. DeepFM can simultaneously capture low-order and high-order interactions among different features and is suitable for handling large-scale sparse data.

4.1.3 Implementation Details. Since different recommendation models have different requirements for input data, we use different data features according to the actual needs of each model for training. The training features for different recommendation models are summarized in Table 2. Besides, all model parameters are set according to the default best values provided in the original paper. After completing the model training, we simulated the behaviors of 151,021 users from MIND's test set and conducted a total of 6 rounds of interaction simulations. We use 101,527 news items from the training set as the news database. In each round of simulation, the number of recalled news is 500, and the number of recommended news for users is 50. The experiments were run on a Linux server with 2 NVIDIA GeForce RTX 2080Ti GPUS, and each with 11GB GPU memory. To make the experiments reproducible, we make the codes publicly available at here.

Table 2: Training features of different models.

Models	Training features
NCF	user ID, news ID, user-news interaction
NGCF	user ID, news ID, user-news interaction
NRMS	user ID, news ID, news title, user-news interaction
NAML	user ID, news ID, news title, news abstract, news topic category and subcategory, user-news interaction
DKN	user ID, news ID, news title, entities in news title, user-news interaction
DeepFM	user ID, news ID, news topic category and subcategory, user-news interaction

6 EVALUATION OF ICCF

6.1 Experimental Setup

6.1.1 Models. The relevant models are summarized as follows:

- DPP [6]: Re-ranking diversity model based on determinant point process.
- MMR [4]: Re-ranking diversity model based on maximal marginal relevance.
- DPP-ICCF: Combined re-ranking model of DPP and ICCF. It selects the current best item from the candidate news list based on the DPP strategy and utilizes ICCF to decide whether to retain it.
- MMR-ICCF: Combined re-ranking model of MMR and ICCF. It is similar to DPP-ICCF, with the main difference being that it selects the current best item from the candidate news list based on the MMR strategy.
- NONE: This model serves as the baseline for ICCF. It does not perform any operation during the re-ranking phase and directly recommends the news with the highest predicted click probability to the users.

6.1.2 Evaluation Metrics. Accuracy evaluation metrics. We use AUC (area under the curve) and NDCG (normalized discounted cumulative gain) to assess the accuracy of the model. AUC is the area under the ROC curve, and a higher AUC value indicates that the model can better distinguish between positive and negative instances. NDCG is used to measure the ranking quality of recommendation systems and is a normalized version of the DCG metric.

Diversity evaluation metrics. We utilize the UAR (user relevance radius) and NCE (normalized category entropy) defined in Section 3.3 to evaluate the diversity of the recommendation results.

Comprehensive evaluation metric. We design a comprehensive evaluation metric (CEM) that considers both accuracy and diversity to assess the overall performance of the model. First, get the average accuracy $\bar{P} = (AUC + NDCG)/2$, and the average diversity $\bar{D} = (UAR + NCE)/2$. Then, $CEM = 2\bar{P}\bar{D}/(\bar{P} + \bar{D})$.

6.1.3 Implementation Details. The dataset and basic experimental settings remain the same as before. We evaluate the accuracy and diversity of the recommendation results using user behavior data from the MIND test set. Different from the above simulation experiments, we need news data with user click labels to test the model's performance. Therefore, limited by the existing data, we only use labeled news as the candidate set, from which news is selected for recommendation. We filtered the user behavior records to ensure that each user has at least 10 news click records and 100 candidate news available for recommendation. Note that due to the limited number of candidate news, the final news for recommendation may not reach the target number after filtering with ICCF. In such cases, we pad the recommendation list with news that have highest predicted click probabilities.

REFERENCES

[1] Ashton Anderson, Lucas Maystre, Ian Anderson, Rishabh Mehrotra, and Mounia Lalmas. 2020. Algorithmic Effects on the Diversity of Consumption on Spotify. In *WWW '20: The Web Conference 2020*. ACM / IW3C2, 2155–2165. https://doi.org/10.1145/3366423.3380281

[2] Guy Aridor, Duarte Gonçalves, and Shan Sikdar. 2020. Deconstructing the Filter Bubble: User Decision-Making and Recommender Systems. In *RecSys 2020: Fourteenth ACM Conference on Recommender Systems*. ACM, 82–91. https://doi.org/10.1145/3383313.3412246

[3] Jesús Bobadilla, Fernando Ortega, Antonio Hernando, and Abraham Gutiérrez. 2013. Recommender systems survey. *Knowl. Based Syst.* 46 (2013), 109–132. https://doi.org/10.1016/j.knosys.2013.03.012

[4] Jaime G. Carbonell and Jade Goldstein. 2017. The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. *SIGIR Forum* 51, 2 (2017), 209–210. https://doi.org/10.1145/3130348.3130369

[5] Allison J. B. Chaney, Brandon M. Stewart, and Barbara E. Engelhardt. 2018. How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. In *Proceedings of the 12th ACM Conference on Recommender Systems*. ACM, 224–232. https://doi.org/10.1145/3240323.3240370

[6] Laming Chen, Guoxin Zhang, and Eric Zhou. 2018. Fast Greedy MAP Inference for Determinantal Point Process to Improve Recommendation Diversity. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018*. 5627–5638.

[7] Tim Donkers and Jürgen Ziegler. 2021. The Dual Echo Chamber: Modeling Social Media Polarization for Interventional Recommending. In *RecSys '21: Fifteenth ACM Conference on Recommender Systems*. ACM, 12–22. https://doi.org/10.1145/3460231.3474261

[8] Seth Flaxman, Sharad Goel, and Justin M Rao. 2016. Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly* 80, S1 (2016), 298–320.

[9] Lu Gan, Diana Nurbakova, Léa Laporte, and Sylvie Calabretto. 2020. Enhancing Recommendation Diversity using Determinantal Point Processes on Knowledge Graphs. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*. ACM, 2001–2004. https://doi.org/10.1145/3397271.3401213

[10] Yingqiang Ge, Shuya Zhao, Honglu Zhou, Changhua Pei, Fei Sun, Wenwu Ou, and Yongfeng Zhang. 2020. Understanding Echo Chambers in E-commerce Recommender Systems. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*. ACM, 2261–2270. https://doi.org/10.1145/3397271.3401431

[11] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. 2017. DeepFM: A Factorization-Machine based Neural Network for CTR Prediction. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*. ijcai.org, 1725–1731. https://doi.org/10.24963/ijcai.2017/239

[12] Han Han, Can Wang, Yunwei Zhao, Min Shu, Wenlei Wang, and Yong Min. 2022. SSLE: A framework for evaluating the "Filter Bubble" effect on the news aggregator and recommenders. *World Wide Web* 25, 3 (2022), 1169–1195. https://doi.org/10.1007/s11280-022-01031-4

[13] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural Collaborative Filtering. In *Proceedings of the 26th International Conference on World Wide Web*. ACM, 173–182. https://doi.org/10.1145/3038912.3052569

[14] Folasade Olubusola Isinkaye, Yetunde O Folajimi, and Bolande Adefowoke Ojokoh. 2015. Recommendation systems: Principles, methods and evaluation. *Egyptian informatics journal* 16, 3 (2015), 261–273.

[15] Ray Jiang, Silvia Chiappa, Tor Lattimore, András György, and Pushmeet Kohli. 2019. Degenerate Feedback Loops in Recommender Systems. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, 383–390. https://doi.org/10.1145/3306618.3314288

[16] Jeff Johnson, Matthijs Douze, and Hervé Jégou. 2021. Billion-Scale Similarity Search with GPUs. *IEEE Trans. Big Data* 7, 3 (2021), 535–547. https://doi.org/10.1109/TBDATA.2019.2921572

[17] Nian Li, Chen Gao, Jinghua Piao, Xin Huang, Aizhen Yue, Liang Zhou, Qingmin Liao, and Yong Li. 2022. An Exploratory Study of Information Cocoon on Short-form Video Platform. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. ACM, 4178–4182. https://doi.org/10.1145/3511808.3557548

[18] Yile Liang, Tieyun Qian, Qing Li, and Hongzhi Yin. 2021. Enhancing Domain-Level and User-Level Adaptivity in Diversified Recommendation. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 747–756. https://doi.org/10.1145/3404835.3462957

[19] Ping Liu, Karthik Shivaram, Aron Culotta, Matthew A. Shapiro, and Mustafa Bilgic. 2021. The Interaction between Political Typology and Filter Bubbles in News Recommendation Algorithms. In *WWW '21: The Web Conference 2021*. ACM / IW3C2, 3791–3801. https://doi.org/10.1145/3442381.3450113

[20] Yujie Lu, Shengyu Zhang, Yingxuan Huang, Luyao Wang, Xinyao Yu, Zhou Zhao, and Fei Wu. 2021. Future-Aware Diverse Trends Framework for Recommendation. In *WWW '21: The Web Conference 2021*. ACM / IW3C2, 2992–3001. https://doi.org/10.1145/3442381.3449791

[21] Masoud Mansoury, Himan Abdollahpouri, Mykola Pechenizkiy, Bamshad Mobasher, and Robin Burke. 2020. Feedback Loop and Bias Amplification in Recommender Systems. In *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management*. ACM, 2145–2148. https://doi.org/10.1145/3340531.3412152

[22] Marco Minici, Federico Cinus, Corrado Monti, Francesco Bonchi, and Giuseppe Manco. 2022. Cascade-based Echo Chamber Detection. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. ACM, 1511–1520. https://doi.org/10.1145/3511808.3557253

[23] Tien T. Nguyen, Pik-Mai Hui, F. Maxwell Harper, Loren G. Terveen, and Joseph A. Konstan. 2014. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *23rd International World Wide Web Conference*. ACM, 677–686. https://doi.org/10.1145/2566486.2568012

[24] Dimitar Nikolov, Mounia Lalmas, Alessandro Flammini, and Filippo Menczer. 2019. Quantifying Biases in Online Information Exposure. *J. Assoc. Inf. Sci. Technol.* 70, 3 (2019), 218–229. https://doi.org/10.1002/asi.24121

[25] Dimitar Nikolov, Diego F. M. Oliveira, Alessandro Flammini, and Filippo Menczer. 2015. Measuring online social bubbles. *PeerJ Comput. Sci.* 1 (2015), e38. https://doi.org/10.7717/peerj-cs.38

[26] Fernando P. Santos, Yphtach Lelkes, and Simon A. Levin. 2021. Link recommendation algorithms and dynamics of polarization in online social networks. *Proc. Natl. Acad. Sci. USA* 118, 50 (2021), e2102141118. https://doi.org/10.1073/pnas.2102141118

[27] Ayan Sinha, David F. Gleich, and Karthik Ramani. 2016. Deconvolving Feedback Loops in Recommender Systems. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016*. 3243–3251.

[28] Jianing Sun, Wei Guo, Dengcheng Zhang, Yingxue Zhang, Florence Regol, Yaochen Hu, Huifeng Guo, Ruiming Tang, Han Yuan, Xiuqiang He, and Mark Coates. 2020. A Framework for Recommending Accurate and Diverse Items Using Bayesian Graph Convolutional Neural Networks. In *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM, 2030–2039. https://doi.org/10.1145/3394486.3403254

[29] Wenlong Sun, Sami Khenissi, Olfa Nasraoui, and Patrick Shafto. 2019. Debiasing the Human-Recommender System Feedback Loop in Collaborative Filtering. In *Companion of The 2019 World Wide Web Conference*. ACM, 645–651. https://doi.org/10.1145/3308560.3317303

[30] Hongwei Wang, Fuzheng Zhang, Xing Xie, and Minyi Guo. 2018. DKN: Deep Knowledge-Aware Network for News Recommendation. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*. ACM, 1835–1844. https://doi.org/10.1145/3178876.3186175

[31] Wenjie Wang, Fuli Feng, Liqiang Nie, and Tat-Seng Chua. 2022. User-controllable Recommendation Against Filter Bubbles. In *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 1251–1261. https://doi.org/10.1145/3477495.3532075

[32] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. 2019. Neural Graph Collaborative Filtering. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 165–174. https://doi.org/10.1145/3331184.3331267

[33] Chuhan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang, and Xing Xie. 2019. Neural News Recommendation with Attentive Multi-View Learning. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*. ijcai.org, 3863–3869. https://doi.org/10.24963/ijcai.2019/536

[34] Chuhan Wu, Fangzhao Wu, Suyu Ge, Tao Qi, Yongfeng Huang, and Xing Xie. 2019. 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*. Association for Computational Linguistics, 6388–6393. https://doi.org/10.18653/v1/D19-1671

[35] Fangzhao Wu, Ying Qiao, Jiun-Hung Chen, Chuhan Wu, Tao Qi, Jianxun Lian, Danyang Liu, Xing Xie, Jianfeng Gao, Winnie Wu, and Ming Zhou. 2020. 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, 3597–3606. https://doi.org/10.18653/v1/2020.acl-main.331

[36] Le Wu, Xiangnan He, Xiang Wang, Kun Zhang, and Meng Wang. 2023. A Survey on Accuracy-Oriented Neural Recommendation: From Collaborative Filtering to Information-Rich Recommendation. *IEEE Trans. Knowl. Data Eng.* 35, 5 (2023), 4425–4445. https://doi.org/10.1109/TKDE.2022.3145690

[37] Rui Ye, Yuqing Hou, Te Lei, Yunxing Zhang, Qing Zhang, Jiale Guo, Huaiwen Wu, and Hengliang Luo. 2021. Dynamic Graph Construction for Improving Diversity of Recommendation. In *RecSys '21: Fifteenth ACM Conference on Recommender Systems*. ACM, 651–655. https://doi.org/10.1145/3460231.3478845

[38] Han Zhang, Ziwei Zhu, and James Caverlee. 2023. Evolution of Filter Bubbles and Polarization in News Recommendation. In *Advances in Information Retrieval - 45th European Conference on Information Retrieval (Lecture Notes in Computer Science, Vol. 13981)*. Springer, 685–693. https://doi.org/10.1007/978-3-031-28238-6_60

[39] Yu Zheng, Chen Gao, Liang Chen, Depeng Jin, and Yong Li. 2021. DGCN: Diversified Recommendation with Graph Convolutional Networks. In *WWW '21: The Web Conference 2021*. ACM / IW3C2, 401–412. https://doi.org/10.1145/3442381.3449835

[40] Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, and Georg Lausen. 2005. Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web*. ACM, 22–32. https://doi.org/10.1145/1060745.1060754