Sequential Recommenders and Multimodal Inputs: Mitigating Data Quality Issues in Industry-Scale Recommenders

Jan Malte Lichtenberg[1,*], Matteo Ruffini[1]

[1]Albatross AI

Abstract

Poor data quality remains a key bottleneck for advancing recommender systems in industrial settings. In this talk, we argue that sequential transformer-based recommenders, particularly ID-based architectures such as SASRec [1], are more robust to several forms of data quality issues compared to traditional learning-to-rank approaches. However, they suffer acutely from item cold-start, which we treat as a missing-modality problem. We then discuss how multimodal content embeddings can address this challenge and present *DenseRec* [2], a simple but effective method to mitigate item cold-start by integrating dense content embeddings into sequential models.

Keywords

Data Quality, Multimodal Recommenders, Sequential Recommenders, Item Cold Start, Content Embeddings

Introduction. Research in recommender systems continues to produce increasingly sophisticated models, yet their practical impact is often constrained by the quality of the underlying data. In industry-scale settings, data quality (DQ) issues such as faulty feature values, feedback attribution errors, and cold-start scenarios can dominate algorithmic considerations [3]. Our talk addresses two central questions: (i) Are some recommender architectures inherently more robust to DQ issues than others? (ii) How can knowledge of such issues inform the design of multimodal sequential recommenders?

Data Quality Challenges in Traditional Approaches. Typical learning-to-rank (LTR) setups rely heavily on hand-crafted features and explicit feedback attribution, both of which are prone to noise and missingness. In practice, we observe five recurring sources of DQ issues: (1) ambiguous or delayed feedback attribution, (2) missing or faulty item and user features, (3) user cold-start, (4) item cold-start, and (5) inaccurate propensity estimates for off-policy evaluation. Addressing these challenges is particularly critical for recommendation-as-a-service providers, where attribution and feature quality issues can compound across customers. The engineering overhead of maintaining feature pipelines and ensuring data quality becomes a significant bottleneck, especially for small teams scaling recommendation systems.

Sequential Recommenders as a Robust Alternative. ID-based sequential recommenders such as SASRec [1] circumvent several of the above challenges. They model user behavior directly from sequences of interactions without requiring additional feedback signals or extensive feature engineering. As a result, many DQ issues are transformed into model or optimization challenges rather than upstream data-engineering problems. This architectural choice provides inherent robustness: noisy features cannot corrupt the model if they are not used, and feedback attribution is simplified to the sequence order itself. However, one major weakness of this class of models lies in item cold-start. Unseen items cannot be recommended, nor can they contribute to updating a user's state. This reduces both reactivity and coverage, creating a form of missing-modality problem where behavioral signals are entirely absent for new items.

Multimodal Inputs for Item Cold Start. A natural solution is to enrich sequential recommenders with content-based embeddings of items (e.g., text descriptions, images, audio). However, recent studies (e.g., [4, 5, 6]) show that naïvely injecting content embeddings often fails to close the cold-start gap and can even harm performance on warm items. The key design challenge is how to combine behavioral ID

DaQuAMRec 2025: 1st International Workshop on Data Quality-Aware Multimodal Recommendation, September 22nd, 2025, Prague, Czech Republic
malte@usealbatross.ai (J. M. Lichtenberg); matteo@usealbatross.ai (M. Ruffini)

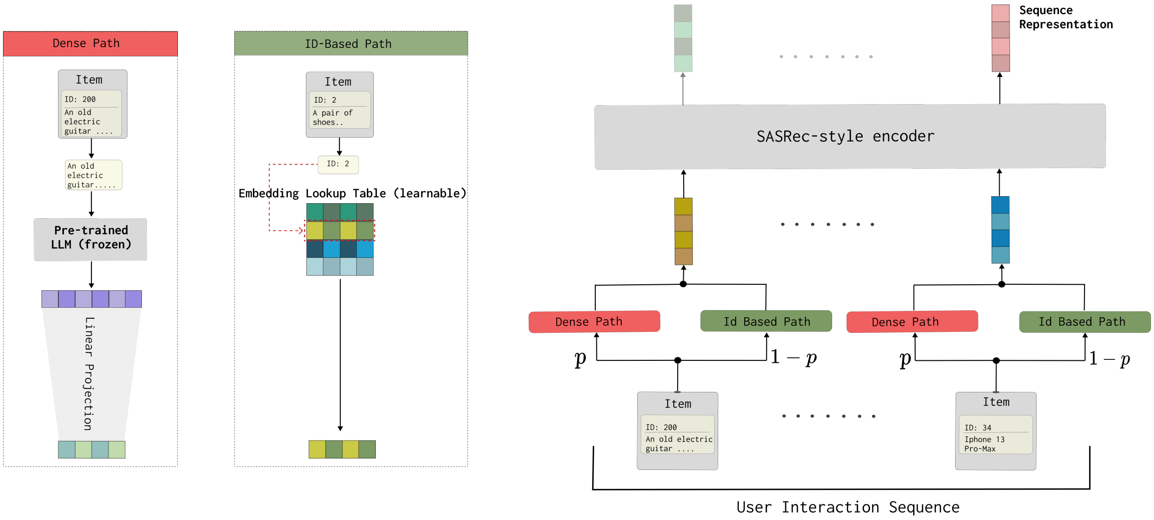

Figure 1: DenseRec [2] architecture overview. The model maintains two parallel embedding pathways: (1) ID Path using traditional learnable embeddings, and (2) Content Path using pre-computed content embeddings projected into the ID embedding space via a learnable projection layer. During training, a probabilistic selection mechanism determines which path to use for each token position.

embeddings with dense multimodal representations in a way that preserves the advantages of both. This connects directly to the workshop's focus on data quality in multimodal recommendation [3]: item cold-start represents a severe form of missing-modality scenario where the behavioral modality is completely absent [7].

DenseRec: A Dual-Path Approach. We proposed *DenseRec* [2], a simple yet effective dual-path method for sequential recommendation. DenseRec (Figure 1) retains a primary ID-based path while adding a secondary "dense path" derived from pretrained content embeddings. A learnable projection layer aligns the dense representations with the ID embedding space. During training, both paths are stochastically activated with probability p_{dense}, a hyper parameter. At inference time, the dense path is used only for cold-start items, while known items rely exclusively on the ID path. This ensures that multimodal signals augment the model *only where necessary*, thereby reducing destructive interference with strong behavioral signals. The approach treats item cold-start as a missing-modality problem and applies a selective integration strategy to maintain recommendation quality across both warm and cold items.

Experimental Validation. We evaluated DenseRec on three Amazon Reviews 2023 datasets [8] using time-based train/validation/test splits with strict cold-start: unseen items appear only in the test set. For dense content embeddings, we used the `all-MiniLM-L6-v2` model[1] from the `sentence-transformers` library [9] to embed item content. We compared DenseRec against a strong ID-only SASRec baseline, applying hyperparameter optimization only to the baseline for fairness. Results, measured in HR@100, show that DenseRec significantly mitigates item cold-start, with performance varying smoothly as a function of p_{dense}. In particular, $p_{\text{dense}} = 0.5$ yielded a favorable balance between ID and dense signals, demonstrating that careful integration of multimodal content can address data quality issues arising from missing behavioral data. In ongoing work, we experiment with multi-modal embeddings to build upon the results presented in [2].

Conclusion. Sequential ID-based recommenders provide robustness against many DQ issues prevalent in traditional LTR approaches, but struggle with item cold-start. DenseRec introduces a lightweight, non-destructive mechanism to incorporate multimodal content embeddings, reducing cold-start while retaining the strengths of ID-based modeling. More broadly, our findings highlight that understanding DQ challenges, including missing modalities in the form of cold-start, can directly inform the design of multimodal sequential models.

[1] https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] W.-C. Kang, J. McAuley, Self-attentive sequential recommendation, in: 2018 IEEE international conference on data mining (ICDM), IEEE, 2018, pp. 197–206.

[2] J. M. Lichtenberg, A. De Candia, M. Ruffini, Denserec: Revisiting dense content embeddings for sequential transformer-based recommendation, arXiv preprint arXiv:2508.18442 (2025).

[3] C. Pomo, D. Jannach, Y. Kim, D. Malitesta, A. C. M. Mancino, J. McAuley, A. Melchiorre, S. Nawaz, First international workshop on data quality-aware multimodal recommendation (daquamrec), in: Proceedings of the Nineteenth ACM Conference on Recommender Systems, 2025, pp. 1378–1382.

[4] A. Singh, T. Vu, N. Mehta, R. Keshavan, M. Sathiamoorthy, Y. Zheng, L. Hong, L. Heldt, L. Wei, D. Tandon, et al., Better generalization with semantic ids: A case study in ranking for recommendations, in: Proceedings of the 18th ACM Conference on Recommender Systems, 2024, pp. 1039–1044.

[5] Y. Hou, Z. He, J. McAuley, W. X. Zhao, Learning vector-quantized item representation for transferable sequential recommenders, in: Proceedings of the ACM Web Conference 2023, 2023, pp. 1162–1171.

[6] L. Zhang, X. Zhou, Z. Zeng, Z. Shen, Are id embeddings necessary? whitening pre-trained text embeddings for effective sequential recommendation, in: 2024 IEEE 40th International Conference on Data Engineering (ICDE), IEEE, 2024, pp. 530–543.

[7] C. Ganhör, M. Moscati, A. Hausberger, S. Nawaz, M. Schedl, A multimodal single-branch embedding network for recommendation in cold-start and missing modality scenarios, in: Proceedings of the 18th ACM Conference on Recommender Systems, 2024, pp. 380–390.

[8] Y. Hou, J. Li, Z. He, A. Yan, X. Chen, J. McAuley, Bridging language and items for retrieval and recommendation, arXiv preprint arXiv:2403.03952 (2024).

[9] N. Reimers, I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2019. URL: https://arxiv.org/abs/1908.10084.