

Mend the Measurement Gap: Latent User Preference Modeling for Short-Form Video Recommendation

Shuo Chang
Google
Mountain View, CA, USA
changshuo@google.com

Ali Montazer
Google
Mountain View, CA, USA
alimontazer@google.com

Dapeng Hong
Google
Mountain View, CA, USA
dapengh@google.com

Yueqi Wang
Google
New York City, NY, USA
yueqiw@google.com

Jiangguo Zhang
Google
Mountain View, CA, USA
jiangguozhang@google.com

Tomer Margolin
Google
Mountain View, CA, USA
tmargolin@google.com

Ningren Han
YouTube
Google
Mountain View, CA, USA
peterhan@google.com

Zihuan Diao
Google
Mountain View, CA, USA
zdiao@google.com

Joyneel Misra
Google
Mountain View, CA, USA
jomi@google.com

Sourabh Bansod
Google
Mountain View, CA, USA
spbansod@google.com

Abstract

Recommender systems rely heavily on heterogeneous behavioral feedback to infer user preference. Although abundant, these signals are imperfect measurements: the same observed behavior can arise from different underlying states, such as genuine enjoyment, passive consumption, or inattention. The challenge is especially acute in short-form video, where watch-based signals are strongly affected by measurement confounders such as video duration — the same watch time can imply different levels of preference for videos of different lengths, while ratio-based metrics can systematically favor short videos. As a result, optimizing raw engagement can amplify measurement artifacts rather than improving user value. We propose a Factorized Latent Value Model (FLVM) for measuring user preference from heterogeneous behavioral feedback. The model treats observed behaviors as noisy measurements of a low-dimensional, factorized latent value state and uses structured output heads to model heterogeneous feedback signals. A restricted baseline path captures predictable variation from measurement-confounding features such as video duration, user propensity, and session context, while a routed latent path estimates preference-relevant value advantage. The resulting latent value score can be integrated into an existing recommender system as a ranking feature or ranking score. On YouTube Shorts, a major short-form video

platform, this model improves offline metrics and lifts a primary viewer enjoyment metric by 2.67% in online A/B tests.

CCS Concepts

• Information systems → Recommender systems; Personalization.

Keywords

preference measurement, latent variable models, feedback debiasing, multi-task recommendation, short-form video recommendation

ACM Reference Format:

Shuo Chang, Yueqi Wang, Zihuan Diao, Ali Montazer, Jiangguo Zhang, Joyneel Misra, Dapeng Hong, Tomer Margolin, Sourabh Bansod, and Ningren Han. 2026. Mend the Measurement Gap: Latent User Preference Modeling for Short-Form Video Recommendation. In *20th ACM Conference on Recommender Systems (RecSys '26)*, September 27–October 02, 2026, Minneapolis, MN, USA. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3773078.3831923>

1 Introduction

Recommender systems cannot observe user preference directly; they can only infer it from behavior. That inference is especially unreliable in a low-friction sequential recommendation feed. In this setting, the user consumes a continuous stream of recommendations through swipes, likes, shares, comments, and creator subscriptions. Since explicit feedback is sparse, production systems rely heavily on implicit engagement signals such as watch time and completion ratio [9, 26].

But behavioral signals are not direct observations of preference [11, 25]; they are jointly shaped by user preference, content,



This work is licensed under a Creative Commons Attribution 4.0 International License. RecSys '26, Minneapolis, MN, USA

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2284-4/26/09

<https://doi.org/10.1145/3773078.3831923>

video duration, session context, and the system policy. The problem is especially acute on such surfaces, in contrast to click-based discovery pages with active selection [9, 19]. The same observed engagement can reflect different underlying states: a user may watch a video for a long time out of genuine enjoyment or mere passive consumption. A short watch may indicate low preference, quick satisfaction, or active skipping in search of something else.

This ambiguity creates a fundamental challenge for ranking optimization. Raw watch time is strongly affected by video duration, while objectives based on the completion ratio (watch time divided by video duration) can introduce the opposite bias toward shorter videos [8, 18, 22, 27]. Explicit actions such as likes, dislikes, and shares are stronger preference indicators but sparse and highly user-dependent. A standard multi-task ranker can learn correlations among these feedback labels [9, 12, 26], yet it still directly optimizes observed labels whose meanings are duration-dependent, user-dependent, and semantically entangled.

We call this the *measurement gap*: the distance between the behaviors a recommender observes and the underlying user preference it is trying to measure. Prior work has shown that passive-negative feedback, survey responses, and explicit negative actions provide useful evidence beyond positive engagement alone [14, 15, 21]. We treat these heterogeneous signals jointly as noisy measurements of a shared latent value state¹, introducing an explicit measurement layer between raw behavioral feedback and ranking optimization.

Building on latent variable recommendation and baseline-relative debiasing, we propose a **Factorized Latent Value Model** (FLVM) as a structured measurement layer for measuring user preference from heterogeneous behavioral feedback. Because observed behaviors are noisy measurements of an unobserved construct, a latent variable model is a natural fit: FLVM uses a supervised Gaussian variational bottleneck [1], building on a line of latent variable approaches to recommendation [2, 6]. Given a user-item impression, the model infers a low-dimensional, factorized latent value from serving-time features and sparsely routes its components to decoders for heterogeneous feedback signals. A restricted baseline path, conditioned only on duration, session context, and user propensity, models the portion of observed engagement predictable from measurement-confounding factors; the latent path then explains residual variation across feedback signals.

This model replaces an existing multi-task ranking model in the pre-scoring stage of our deployment; the resulting latents serve as primary inputs to the scoring formula.

The main contributions are:

- (1) We frame preference learning in recommender systems as a measurement problem, where heterogeneous behavioral signals are noisy and confounded observations of latent user preference.

- (2) We introduce FLVM, a structured latent measurement layer that combines a restricted baseline over measurement-confounding features with sparsely routed latent decoders over heterogeneous feedback signals.
- (3) We instantiate FLVM in a production short-form video recommender, using duration, session context, and user propensity as baseline features and watch, engagement, valence, and satisfaction-like labels as feedback heads.
- (4) We validate FLVM offline and online on YouTube Shorts, a major short-form video platform, showing improved prediction of key feedback signals and a 2.67% lift in a primary viewer enjoyment metric with neutral guardrails, with consistent gains carrying over to newly uploaded content.

2 Related Work

Behavior as a noisy measure of preference. Recent industry work distinguishes observed engagement from the preference or satisfaction a recommender seeks to optimize. In short-form video feeds, peak-end-based user-return modeling shows that impressions should not contribute equally to retention [19], while direct survey supervision can reveal interest that watch time, likes, and shares miss [11]. Survey labels are themselves sparse and selected: response-propensity modeling improves satisfaction estimation under survey response bias [3], and multi-head survey models capture satisfaction and negative experiences in production [21]. Skips and explicit negative-feedback controls provide complementary evidence beyond positive engagement [14, 15]. FLVM shares this measurement motivation but treats these impression-level signals jointly as noisy indicators of a structured latent state, rather than optimizing a single survey or retention target.

Watch-time and feedback debiasing. Although widely used as a supervision signal, watch time is shaped by video duration and does not always reflect user interest. Prior work uses duration-stratified or causal labels [22, 23], separates interest-driven watch time from time spent before users recognize disinterest [25], and develops architectural and counterfactual methods that suppress duration information or separate its effects [18, 24, 27]. CVRDD removes duration’s direct effect through counterfactual inference, whereas the later CWM formulation treats observed watch time as a duration-truncated realization of latent counterfactual watch time. Most closely related to our baseline-relative construction, Relative Advantage Debiasing maps watch time against user- and item-conditioned reference distributions to create a relative preference target [8]. AlignPxtr aligns continuous and discrete behavior distributions across multiple bias dimensions [7]. Bucketed and prototype-based models further motivate structured watch-time outputs [4, 17]. FLVM instead uses a restricted baseline over duration, session context, and user propensity, and learns a residual variational latent jointly across heterogeneous feedback signals.

Variational and multi-task recommendation. FLVM’s stochastic latent layer follows the supervised variational information bottleneck formulation [1]. In recommendation, Multi-VAE adapts variational inference to implicit collaborative filtering, while MacridVAE learns disentangled intent representations [6, 13]. Multi-Behavior Alignment treats universal user preference as a latent variable inferred

¹Here, “value” denotes the latent construct that observed engagement signals imperfectly measure—not the reinforcement-learning value function or long-term user value.

from several behavior distributions, and BVAE combines behavior-aware variational representations with multi-task outputs [16, 20]. Chang et al. [2] provide a production precedent for injecting VAE-inferred user intent into a sequential recommender. Production rankers otherwise commonly predict several feedback objectives directly using shared multi-task architectures [9, 12, 26]. FLVM differs by learning an impression-level, semantically routed bottleneck from heterogeneous measurements and pairing it with a stopped, restricted baseline before downstream scoring.

3 Factorized Latent Value Model

3.1 Problem Formulation

The measurement gap motivates treating observed behaviors as noisy measurements rather than direct preference labels. We call the unobserved construct underlying these behaviors *latent value*; because it carries no label of its own, the model must infer it rather than optimize it directly. It is impression-level, capturing the preference-relevant signal for a specific user-item pairing in context.

For each impression, let x denote the serving-time feature vector and let $x_b \subseteq x$ denote a subset of measurement-confounding features. In our implementation, x_b includes video duration, user propensity, and session context. Let

$$y = \{y_1, y_2, \dots, y_K\} \quad (1)$$

denote the K heterogeneous feedback signals observed after the impression, such as watch-time buckets, completion, likes, shares, negative actions, and survey-style feedback.

As shown in Figure 1, the model has the dependency structure

$$x \rightarrow z, \quad (x_b, z) \rightarrow y \quad (2)$$

The encoder defines a conditional distribution $q_\phi(z | x)$ over the latent state. Given z and x_b , the K feedback heads are conditionally independent, and each conditional $p(y_k | z, x_b)$ is parameterized by combining two structural paths: a *restricted baseline path* b_k given only the confounding features x_b , and a *latent path* h_k given the latent state z . This split is the architectural inductive bias that routes predictable measurement-driven variation to the baseline and lets z carry preference-relevant residual; the two contributions combine in logit space and pass through a task-specific output transformation o_k :

$$\ell_k = f_k(b_k(x_b), h_k(z)), \quad \hat{y}_k = o_k(\ell_k) \quad (3)$$

where o_k is a sigmoid for binary feedback labels and a softmax for categorical labels. The encoder sees x in full, so z can in principle encode any function of x including x_b ; the training scheme encourages the latent path to focus on variation in y_k that the baseline cannot already explain, so z carries residual variation in practice. The next sections specialize f_k and h_k , introduce the structure of z (factorization and routing), and discuss identifiability, training, and serving.

3.2 Baseline and Latent Value Advantage

Our feedback targets are binary or bucketed classification labels, so each decoder operates in logit space. The baseline path is implemented as an MLP; the latent path applies a strictly positive gain to the routed latent state.

The restricted baseline path produces a baseline logit

$$\ell_k^b = b_k(x_b) \quad (4)$$

Because x_b excludes user-video matching features, the baseline has no direct path to user-item relevance.

The latent path produces a logit-scale **latent value advantage**

$$\delta_k = \gamma_k^+ \cdot h_k(z) \quad (5)$$

where h_k depends only on the components of z routed to task k (Section 3.3) and $\gamma_k^+ \equiv \text{softplus}(\alpha_k) > 0$ is a learned per-signal gain. This positivity stabilizes the sign conventions of z used for identifiability (Section 3.4).

The final decoder logit is

$$\ell_k = \text{sg}(\ell_k^b) + \delta_k \quad (6)$$

where $\text{sg}(\cdot)$ denotes the stop-gradient operator. This baseline-relative formulation is conceptually related to relative-advantage approaches for watch-time debiasing [8], but applies the advantage idea to residuals across heterogeneous feedback signals.

Because gradients are stopped through ℓ_k^b , the main feedback objective does not backpropagate into the baseline path: the baseline is updated exclusively from gradients of $\mathcal{L}_{\text{base}}$ and the main loss rewards δ_k only for the portion of y_k not already explained by the baseline. This encourages confounding effects to be captured by the baseline, yielding a debiased residual.

An alternative is a decoder $p(y_k | z, x_b)$ conditioned directly on the baseline features. Without the stop-gradient and auxiliary baseline loss, the decoder can freely mix x_b and z , giving z no reason to encode only residual variation.

3.3 Factorized Latent Value Variables

We factorize the latent value into three components capturing consumption, active engagement, and valence:

$$z = [z_p, z_a, z_s] \quad (7)$$

where z_p captures consumption-related evidence, z_a captures active response propensity, and z_s captures the valence of that engagement: the sentiment-like positive or negative polarity expressed through feedback labels. The valence latent z_s enters jointly with z_a to explain feedback such as deep engagement, negative feedback, and satisfaction surveys, where the same active behavior can reflect either a positive or negative reaction.

This factorization reflects the observation that different user feedback signals tell different stories about the same view. A user might watch a long video to completion and scroll on; another might watch and like it; a third might watch and hit the dislike button. On consumption alone, these views are indistinguishable; only the direction of engagement separates the second from the third, and only whether the user actively responded separates the first from the rest. A single unconstrained latent forces the model to average across these qualitatively different signals.

To align each decoder with semantically relevant parts of the latent state, we restrict its latent path with a per-signal hard sparse routing mask m_k ,

$$\delta_k = \gamma_k^+ \cdot h_k(m_k \odot z) \quad (8)$$

The mask limits each feedback decoder to a small subset of the factorized latent value state, letting related feedback tasks share

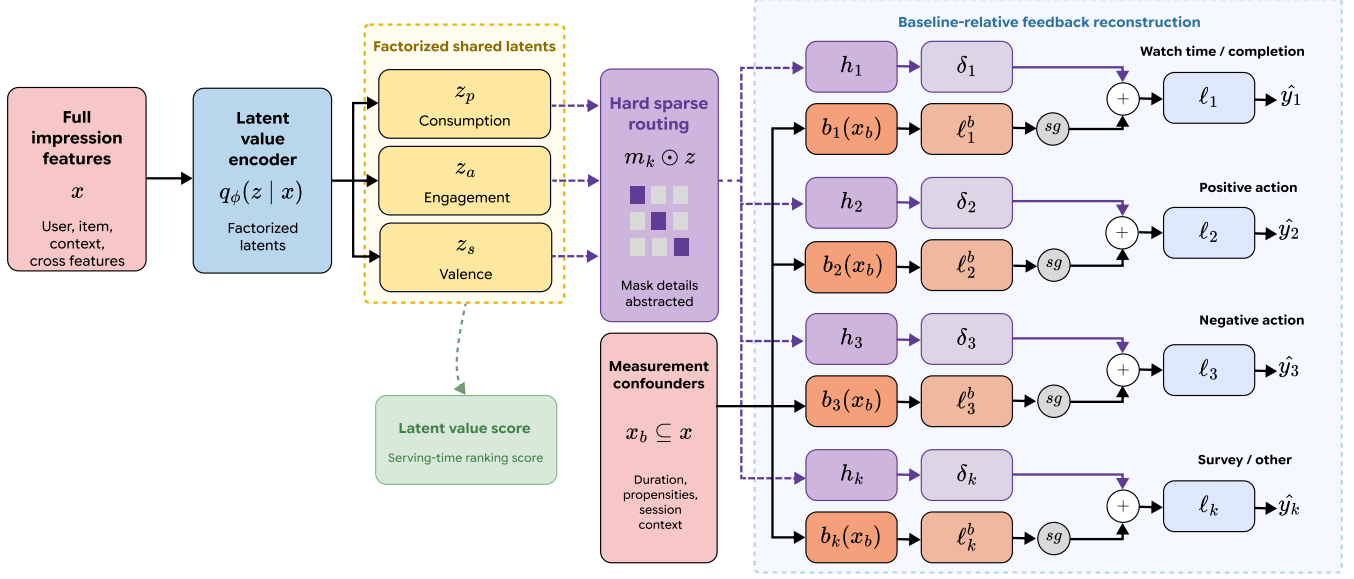


Figure 1: Factorized Latent Value Model. The encoder maps full impression features x to a factorized latent value state $z = [z_p, z_a, z_s]$. For each feedback reconstruction task k , a task-specific hard sparse mask m_k restricts the latent input to decoder h_k , which produces a logit-scale advantage δ_k . In parallel, a restricted baseline b_k uses only measurement-confounding features $x_b \subseteq x$ to produce ℓ_k^b . The final logit is $\ell_k = \text{sg}(\ell_k^b) + \delta_k$, and predictions are $\hat{y}_k = o_k(\ell_k)$. The exact mask configuration is abstracted; the latent value score is scalarized from z for serving-time ranking.

explanatory factors while preventing unrelated tasks from freely using all latent dimensions. Routing follows the factorization: consumption-related labels are explained through z_p , while explicit feedback is explained through z_a and z_s jointly.

This structural constraint also gives the latent components more stable semantics: in an unconstrained bottleneck, latent dimensions can rotate or mix without changing prediction quality, whereas here the meaning of each component is induced by the semantic groups of decoders it is allowed to influence.

3.4 Inductive Bias for Semantic Roles

The semantic labels we attach to z_p , z_a , and z_s are not enforced by the encoder or the standard normal prior alone. A factorized Gaussian latent with an isotropic prior is rotation- and sign-symmetric: nothing in the encoder or the regularization objective prevents the model from arriving at an arbitrary rotation of the three latent axes, in which case the labels “consumption,” “engagement,” and “valence” would be meaningless.

Three architectural choices together reduce this symmetry. First, the routing masks m_k give each latent influence over a distinct set of decoders. A rotation that mixes z_p into z_s would misroute consumption variation into decoders designed to predict valence-signed feedback, increasing the loss. Second, the latent path scales each routed factor by the strictly positive γ_k^+ , so within the set of decoders sharing a factor, the factor’s sign cannot be inverted without loss. Third, the polarity of valence-signed feedback labels, such as explicit satisfaction surveys and negative-engagement signals, anchors the sign convention of z_s . Watch-time label monotonicity

and binary engagement intensities play analogous roles for z_p and z_a .

Together, these three mechanisms supply the inductive bias that encourages stable semantic roles for z_p , z_a , and z_s . Such biases are necessary: Locatello et al. [10] show that without structural assumptions on model or data, disentangled representations cannot be recovered unsupervised. Our architecture provides this bias through decoder assignment and sign-anchored labels, rather than through distributional assumptions. Empirically, liked impressions and impressions carrying explicit negative feedback are most separated along the valence factor z_s , with the two groups 1.7–2.0 standard deviations apart across three models independently trained from scratch. This alignment is stable across training: the sign and the ordering of the three factors are identical in every run, with neutral topline metrics in online A/A testing.

3.5 Training Objective

We train the model with three components: a main feedback loss, an auxiliary baseline anchoring loss, and a KL regularization loss. We use a supervised Gaussian variational bottleneck with reparameterized sampling and KL regularization [1, 5]. The encoder is an MLP with a shared trunk and separate output projections producing a diagonal Gaussian latent value distribution from serving-time features:

$$q_\phi(z | x) = \mathcal{N}(\mu_\phi(x), \text{diag}(\sigma_\phi^2(x))) \quad (9)$$

regularized toward a standard normal prior:

$$p(z) = \mathcal{N}(0, I) \quad (10)$$

Main feedback loss. The main loss trains the encoder and routed latent decoders to predict post-impression feedback:

$$\mathcal{L}_{\text{main}} = \mathbb{E}_{z \sim q_\phi(z|x)} \left[\sum_{k=1}^K r_k \lambda_k \text{CE}_k(y_k, \hat{y}_k) \right] \quad (11)$$

where $r_k \in \{0, 1\}$ is an observation mask indicating whether feedback label k is valid for the impression, and λ_k is the task weight balancing the heterogeneous signals. This is important for sparse feedback, which should not be treated as implicit negatives when missing. Since $\ell_k = \text{sg}(\ell_k^b) + \delta_k$, this loss updates the encoder and routed latent decoders, but not the baseline path.

Auxiliary baseline anchoring loss. The baseline path is trained separately with the same observed labels, but only on x_b :

$$\mathcal{L}_{\text{base}} = \sum_{k=1}^K r_k \lambda_k \text{CE}_k(y_k, \hat{y}_k^b) \quad (12)$$

where $\hat{y}_k^b = o_k(\ell_k^b)$ is the baseline prediction and λ_k is the per-task weight.

Latent regularization. We regularize the latent value distribution with

$$\mathcal{L}_{\text{KL}} = D_{\text{KL}}(q_\phi(z|x) \parallel \mathcal{N}(0, I)) \quad (13)$$

This term keeps the latent space centered and calibrated, reducing the risk that the latent variables become arbitrary task-specific hidden units.

The final training objective is

$$\mathcal{L} = \mathcal{L}_{\text{main}} + \mathcal{L}_{\text{base}} + \beta \mathcal{L}_{\text{KL}} \quad (14)$$

3.6 Ranking Integration

At serving time, we scalarize the latent z into a per-item latent value score using $\sigma(z_s)$ and $\text{softplus}(z_p)$. We omit z_a because it captures active response propensity, while z_s carries the direction of that response. Given that propensity without polarity is not evidence of preference, z_a serves only as a training-time auxiliary construct. The resulting score is a primary input to our ranking formula, replacing the existing multi-task pre-scoring stage.

4 Experiments

We evaluate the model on YouTube Shorts, a major short-form video platform with a billion-scale active user base. We compare against the production pre-scoring ranking model, trained with standard multi-task objectives and a post-hoc score combination stage. Figure 2 shows offline validation of the model’s two main design choices.

Table 1 reports offline metrics across core feedback signals. Skip is omitted from the cross-system comparison because the two systems use different skip-label definitions and are not directly comparable. FLVM reduces item watch-time RMSE and improves Like PR-AUC and Dislike PR-AUC, while Completion sees a slight overall drop. However, a fine-grained duration-sliced analysis across all metrics reveals a consistent debiasing effect: FLVM systematically outperforms the baseline on longer videos ($>60s$) across the board—improving Completion PR-AUC by an average of 0.0032—while scoring lower on shorter videos ($<60s$) by an average of 0.0053. Since the production baseline often inflates

Table 1: Offline predictive performance versus the production multi-task baseline. PR-AUC (higher is better) for binary signals; RMSE (lower is better) for item watch time.

Signal	Metric	Production	FLVM
Item watch time	RMSE ↓	24.47	23.35
Completion	PR-AUC ↑	0.6898	0.6722
Like	PR-AUC ↑	0.2522	0.3190
Dislike	PR-AUC ↑	0.0038	0.0054

Table 2: Decomposition: PR-AUC of each path in FLVM. Baseline-only = $o_k(\ell_k^b)$; Latent-only = $o_k(\delta_k)$; Full = $o_k(\ell_k)$. The two paths together exceed either alone.

Signal	Baseline	Latent	Full
Completion	0.3065	0.6617	0.6722
Like	0.1860	0.0715	0.3190
Skip	0.1840	0.5016	0.5100
Dislike	0.0035	0.0035	0.0054

overall metrics by exploiting short-video heuristics (e.g., inherently higher completion rates), FLVM’s performance shift toward harder-to-consume long content indicates a much better capture of genuine user preference. These improvements suggest that jointly explaining heterogeneous feedback through a compact latent value state denoises sparse and confounded signals more effectively than per-task heads.

Table 2 decomposes FLVM into its two paths. For all signals, the Full model exceeds either path alone, confirming that the baseline and latent contributions are complementary. However, the exact dependency varies by signal and item lifecycle. Notably, we observe a powerful synergistic effect on active engagement (Like): each path predicts Like poorly on its own, yet together they far outperform either alone. This suggests that explicit actions such as Like are heavily masked by measurement-confounding features (e.g., user propensity), and that accounting for these confounders in the restricted baseline lets the latent path recover the preference-relevant signal. For passive behaviors like Completion and Skip, the latent path drives the majority of the predictive performance, though the baseline remains strictly necessary for full accuracy.

In our online A/B experiment, spanning 14 days at our standard experiment traffic allocation, the latent value model improves over our production baseline multi-task pre-scorer, increasing our primary viewer enjoyment metric by **2.67%** with neutral guardrail metrics. To verify the robustness and generalizability of the learned latent value state across different item lifecycles, we evaluated performance on the cold-start slice (fresh videos with limited exposure history). On this group, the model achieved a consistent $>2\%$ lift in helping new videos successfully attract early engagement and transition into mainstream distribution. This demonstrates that

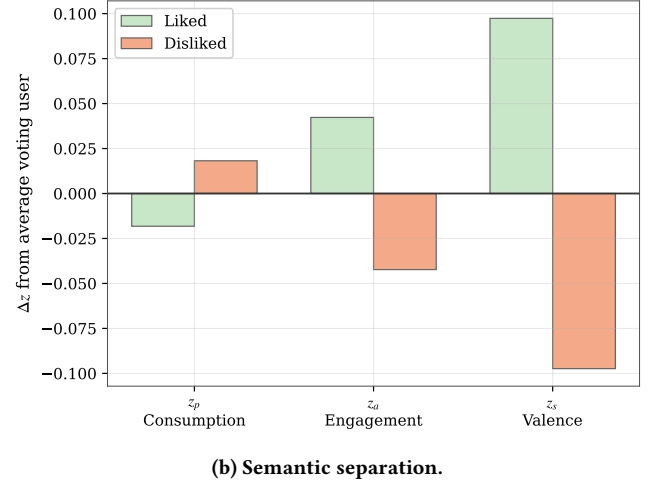
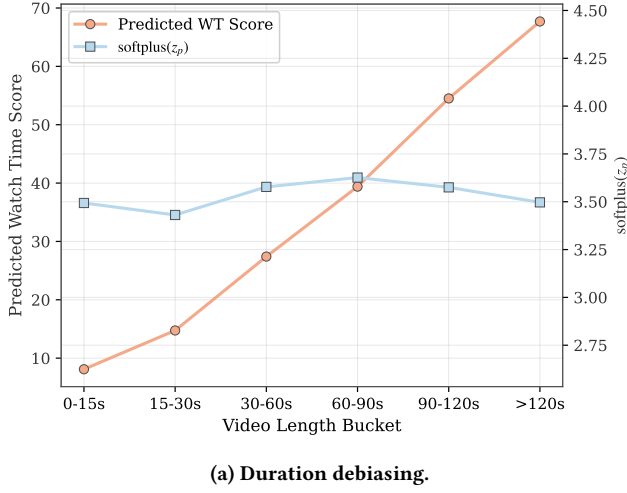


Figure 2: Empirical validation of the two main design choices. (a) The baseline path absorbs duration confounding: predicted item watch time increases with video duration, while the served consumption score $\text{softplus}(z_p)$ stays roughly flat across duration buckets. (b) The three latent factors separate semantically: liked and disliked impressions are most separated along the valence factor z_s , less so along the active-engagement factor z_a , and approximately invariant on the consumption factor z_p .

by separating genuine preference from structural and statistical noise, FLVM avoids overfitting to historic engagement and robustly generalizes to trailing-edge data distributions.

5 Conclusion

We presented a Factorized Latent Value Model that treats observed user feedback as noisy measurements of a low-dimensional, factorized latent value state. The model combines three design choices:

- (1) A baseline restricted to measurement-confounding features.
- (2) Factorized latent value variables that represent different aspects of user preference.
- (3) Predefined hard sparse decoder routing that connects each latent factor only to semantically relevant feedback signals.

The resulting latent value score integrates into an industrial recommender system with meaningful offline and online gains, demonstrating robust generalizability across diverse content lifecycles and varying levels of data sparsity. Empirically, the learned latent factors take on stable semantic roles induced by model structure and feedback signals.

References

- [1] Alexander A. Alemi, Ian Fischer, Joshua V. Dillon, and Kevin Murphy. 2017. Deep Variational Information Bottleneck. In *5th International Conference on Learning Representations, ICLR 2017, Conference Track Proceedings*. OpenReview.net, Toulon, France, 14 pages. arXiv:1612.00410
- [2] Bo Chang, Alexandros Karatzoglou, Yuyan Wang, Can Xu, Ed H. Chi, and Minmin Chen. 2023. Latent User Intent Modeling for Sequential Recommenders. In *Companion Proceedings of the ACM Web Conference 2023 (WWW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 427–431. arXiv:2211.09832 doi:10.1145/3543873.3584641
- [3] Konstantina Christakopoulou, Madeleine Traverse, Trevor Potter, Emma Marriott, Daniel Li, Chris Haulk, Ed H. Chi, and Minmin Chen. 2020. Deconfounding User Satisfaction Estimation from Response Rate Bias. In *Proceedings of the 14th ACM Conference on Recommender Systems (RecSys '20)*. Association for Computing Machinery, New York, NY, USA, 450–455. doi:10.1145/3383313.3412208
- [4] Chao Cui, Shisong Tang, Fan Li, Jiechao Gao, and Hechang Chen. 2025. Calibrating Video Watch-time Predictions with Credible Prototype Alignment. In *Proceedings of the 42nd International Conference on Machine Learning*, Vol. 267. PMLR, Vancouver, Canada, 11563–11584.
- [5] Diederik P. Kingma and Max Welling. 2014. Auto-Encoding Variational Bayes. In *2nd International Conference on Learning Representations, ICLR 2014, Conference Track Proceedings*. OpenReview.net, Banff, AB, Canada, 14 pages. arXiv:1312.6114
- [6] Dawen Liang, Rahul G. Krishnan, Matthew D. Hoffman, and Tony Jebara. 2018. Variational Autoencoders for Collaborative Filtering. In *Proceedings of The Web Conference 2018 (WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 689–698. arXiv:1802.05814 doi:10.1145/3178876.3186150
- [7] Chengzhi Lin, Chuyuan Wang, Annan Xie, Wuhong Wang, Ziyue Zhang, Canguang Ruan, Yuancai Huang, and Yongqi Liu. 2025. AlignPxxr: Aligning Predicted Behavior Distributions for Bias-Free Video Recommendations. arXiv:2503.06920 [cs.LR]
- [8] Emily Liu, Kuan Han, Minfeng Zhan, Bochong Zhao, Guanyu Mu, and Yang Song. 2026. Relative Advantage Debiasing for Watch-Time Prediction in Short-Video Recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence* 40, 18 (2026), 15296–15305. arXiv:2508.11086 doi:10.1609/aaai.v40i18.38555
- [9] Qingyun Liu, Zhe Zhao, Liang Liu, Zhen Zhang, Junjie Shan, Yueneng Li, Shuchao Bi, Lichan Hong, and Ed H. Chi. 2023. Multitask Ranking System for Immersive Feed and No More Clicks: A Case Study of Short-Form Video Recommendation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23)*. Association for Computing Machinery, New York, NY, USA, 4709–4716. doi:10.1145/3583780.3615489
- [10] Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Rätsch, Sylvain Gelly, Bernhard Schölkopf, and Olivier Bachem. 2019. Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations. In *Proceedings of the 36th International Conference on Machine Learning (ICML '19, Vol. 97)*. PMLR, Long Beach, California, USA, 4114–4124. arXiv:1811.12359
- [11] Mengxi Lv, Drew Hogg, Thomas Grubb, Shashank Bassi, Min Li, Cayman Simpson, and Senthil Rajagopalan. 2025. Improve the Personalization of Large-Scale Ranking Systems by Integrating User Survey Feedback. In *Proceedings of the Nineteenth ACM Conference on Recommender Systems (RecSys '25)*. Association for Computing Machinery, New York, NY, USA, 971–974. doi:10.1145/3705328.3748119
- [12] Jiaqi Ma, Zhe Zhao, Xinyang Yi, Jilin Chen, Lichan Hong, and Ed H. Chi. 2018. Modeling Task Relationships in Multi-task Learning with Multi-gate Mixture-of-Experts. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18)*. Association for Computing Machinery, New York, NY, USA, 1930–1939. doi:10.1145/3219819.3220007
- [13] Jianxin Ma, Chang Zhou, Peng Cui, Hongxia Yang, and Wenwu Zhu. 2019. Learning Disentangled Representations for Recommendation. In *Advances in Neural Information Processing Systems* 32. Curran Associates, Inc., Red Hook, NY, USA, 5712–5723. arXiv:1910.14238
- [14] Yunzhu Pan, Chen Gao, Jianxin Chang, Yanan Niu, Yang Song, Kun Gai, Depeng Jin, and Yong Li. 2023. Understanding and Modeling Passive-Negative Feedback

- for Short-Video Sequential Recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems (RecSys '23)*. Association for Computing Machinery, New York, NY, USA, 540–550. doi:10.1145/3604915.3608814
- [15] Madhura Raju, Manisha Sharma, Hongyu Xiong, Bingfeng Deng, and Meng Na. 2025. Leveraging Explicit Negative Feedback in Large-Scale Recommendation Systems: A Case Study. In *Proceedings of the Nineteenth ACM Conference on Recommender Systems (RecSys '25)*. Association for Computing Machinery, New York, NY, USA, 999–1001. doi:10.1145/3705328.3748145
- [16] Qianzhen Rao, Yang Liu, Weike Pan, and Zhong Ming. 2023. BVAE: Behavior-aware Variational Autoencoder for Multi-behavior Multi-task Recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems (RecSys '23)*. Association for Computing Machinery, New York, NY, USA, 625–636. doi:10.1145/3604915.3608781
- [17] Jie Sun, Zhaoying Ding, Xiaoshuang Chen, Qi Chen, Yincheng Wang, Kaiqiao Zhan, and Ben Wang. 2024. CREAD: A Classification-Restoration Framework with Error Adaptive Discretization for Watch Time Prediction in Video Recommender Systems. *Proceedings of the AAAI Conference on Artificial Intelligence* 38, 8 (2024), 9027–9034. arXiv:2401.07521 doi:10.1609/aaai.v38i8.28752
- [18] Shisong Tang, Qing Li, Dingmin Wang, Ci Gao, Wentao Xiao, Dan Zhao, Yong Jiang, Qian Ma, and Aoyang Zhang. 2023. Counterfactual Video Recommendation for Duration Debiasing. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23)*. Association for Computing Machinery, New York, NY, USA, 4894–4903. doi:10.1145/3580305.3599797
- [19] Yuyan Wang, Jing Zhong, Yuxin Cui, Zhaohui Guo, Chuanqi Wei, Yanchen Wang, and Zellux Wang. 2025. Not All Impressions Are Created Equal: Psychology-Informed Retention Optimization for Short-Form Video Recommendation. In *Proceedings of the Nineteenth ACM Conference on Recommender Systems (RecSys '25)*. Association for Computing Machinery, New York, NY, USA, 1022–1025. doi:10.1145/3705328.3748122
- [20] Xin Xin, Xiangyuan Liu, Hanbing Wang, Pengjie Ren, Zhumin Chen, Jiahuan Lei, Xinlei Shi, Hengliang Luo, Joemon M. Jose, Maarten de Rijke, and Zhaochun Ren. 2023. Improving Implicit Feedback-Based Recommendation through Multi-Behavior Alignment. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '23)*. Association for Computing Machinery, New York, NY, USA, 932–941. arXiv:2305.05585 doi:10.1145/3539618.3591697
- [21] Chenghui Yu, Haoze Wu, Jian Ding, Bingfeng Deng, and Hongyu Xiong. 2025. Unified Survey Modeling to Limit Negative User Experiences in Recommendation Systems. In *Proceedings of the Nineteenth ACM Conference on Recommender Systems (RecSys '25)*. Association for Computing Machinery, New York, NY, USA, 1104–1107. doi:10.1145/3705328.3748108
- [22] Ruohan Zhan, Changhua Pei, Qiang Su, Jianfeng Wen, Xueliang Wang, Guanyu Mu, Dong Zheng, Peng Jiang, and Kun Gai. 2022. Deconfounding Duration Bias in Watch-Time Prediction for Video Recommendation. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '22)*. Association for Computing Machinery, New York, NY, USA, 4472–4481. arXiv:2206.06003 doi:10.1145/3534678.3539092
- [23] Yang Zhang, Yimeng Bai, Jianxin Chang, Xiaoxue Zang, Song Lu, Jing Lu, Fuli Feng, Yanan Niu, and Yang Song. 2023. Leveraging Watch-Time Feedback for Short-Video Recommendations: A Causal Labeling Framework. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23)*. Association for Computing Machinery, New York, NY, USA, 4952–4959. arXiv:2306.17426 doi:10.1145/3583780.3615483
- [24] Haiyuan Zhao, Guohao Cai, Jieming Zhu, Zhenhua Dong, Jun Xu, and Ji-Rong Wen. 2024. Counteracting Duration Bias in Video Recommendation via Counterfactual Watch Time. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24)*. Association for Computing Machinery, New York, NY, USA, 4455–4466. arXiv:2406.07932 doi:10.1145/3637528.3671817
- [25] Haiyuan Zhao, Lei Zhang, Jun Xu, Guohao Cai, Zhenhua Dong, and Ji-Rong Wen. 2023. Uncovering User Interest from Biased and Noised Watch Time in Video Recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems (RecSys '23)*. Association for Computing Machinery, New York, NY, USA, 528–539. arXiv:2308.08120 doi:10.1145/3604915.3608797
- [26] Zhe Zhao, Lichan Hong, Li Wei, Jilin Chen, Aniruddh Nath, Shawn Andrews, Aditee Kumthekar, Maheswaran Sathiamoorthy, Xinyang Yi, and Ed H. Chi. 2019. Recommending What Video to Watch Next: A Multitask Ranking System. In *Proceedings of the 13th ACM Conference on Recommender Systems (RecSys '19)*. Association for Computing Machinery, New York, NY, USA, 43–51. doi:10.1145/3298689.3346997
- [27] Yu Zheng, Chen Gao, Jingtao Ding, Lingling Yi, Depeng Jin, Yong Li, and Meng Wang. 2022. DVR: Micro-Video Recommendation Optimizing Watch-Time-Gain under Duration Bias. In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22)*. Association for Computing Machinery, New York, NY, USA, 334–345. doi:10.1145/3503161.3548428