Beyond Black-Box: Structuring Multi-Stage Recommender Systems Using Predicted Intents

Yuyan Wang[1], Cheenar Banerjee[2], Samer Chucri[2], Minmin Chen[3]
[1]Stanford University, [2]Google, Inc. [3]Google DeepMind

Abstract

Modern recommender systems are powered by large-scale ML models trained to predict the next item a user will consume. While effective in many applications, these black-box systems are not able to capture the underlying data-generating process, limiting their ability to generalize in dynamic, high-dimensional environments, especially when consumer choices shift due to unobserved contextual factors. A growing body of work advocates for incorporating consumer intent into recommendation models, but a key challenge remains: intent—defined as a consumer's preference or receptiveness toward content categories—is typically not observable for homepage recommendations, which must be generated *before* any user query or interaction.

We introduce the **Intent-Structured Whole-Page Recommender System (ISRec)**, a general framework that that incorporates *predicted* intents—defined as predicted receptiveness towards different content categories—into a multi-stage recommender system, *without* requiring additional data or explicit labels. ISRec consists of three stages: it first predicts a real-time, personalized distribution over possible intents using past behavioral data; then incorporates these predictions into reward modeling; and finally diversifies the final recommendation list to reflect the full spectrum of inferred intents.

We validate ISRec on YouTube, the world's largest video recommendation platform serving billions of users daily. Large-scale field experiments show that ISRec improves key business outcomes, including a 0.05% increase in daily active users and a 0.09% improvement in overall user enjoyment. Our work demonstrates that, contrary to the common belief that structure limits model flexibility,imposing a structure that is aligned with the DGP can, in fact, improve the performance of black-box ML systems.

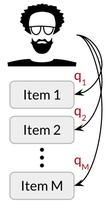
Proposed Framework: Intent-Structured Whole-Page Recommender System (ISRec)

Problem definition: Intent-based recommendation

Expected reward from the set of recommendations R, considering possible **intents:**

$$R^* = \arg\max_R \left[\sum_{v \in \mathcal{V}} \mathbf{P}(v|i,z)\mathbf{P}(R|i,v,z) r_{\mathcal{V}}(j_R|i,z) \right]$$

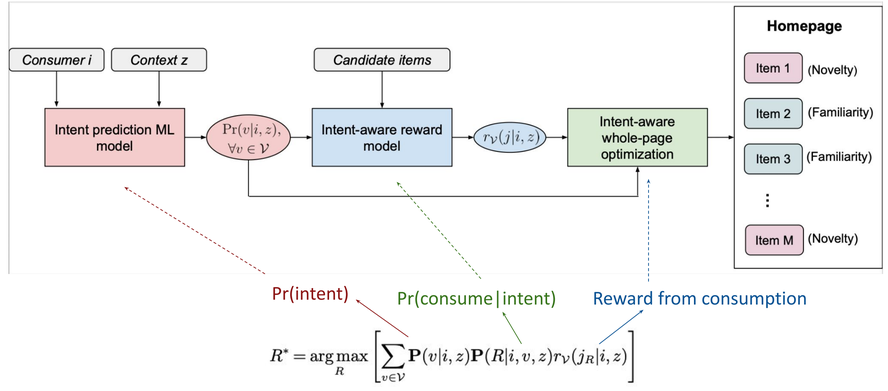
Pr(intent) Pr(consume|intent) Reward from consumption



Notation	Meaning
i	Index on consumer
$v \in \mathcal{V}$	Index on intent
z	Context
j	Index on item
R	Recommendation list
j_R	Item consumed from R

Theorem 1 *The optimization problem above is NP-hard* ***when considering sequential browsing behavior****.*

We will show that the three components actually imply an end-to-end **intent-based recommendation framework**…
… as a greedy solution for the original NP-hard problem!



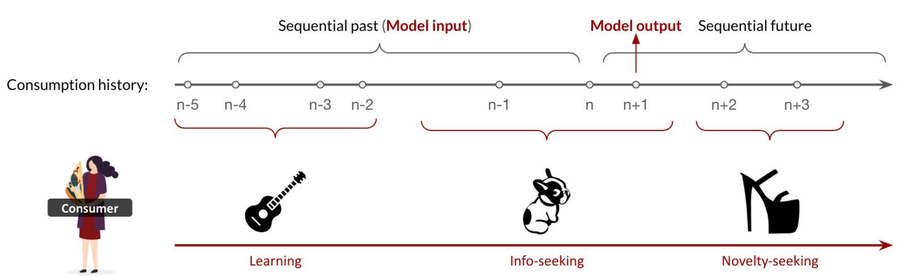
Theorem 2 *The ISRec framework is the optimal greedy solution for the original NP-hard problem, with optimal regret bound (1-1/e).*

The "Data-Generating Process" for Recsys

Existing recommender systems are built on item-level interactions and predicting the next item to be consumed…

However, extrapolating from one item to another can be challenging.

The observation that motivates this work is that consumer behavior are largely driven by their underlying intents:

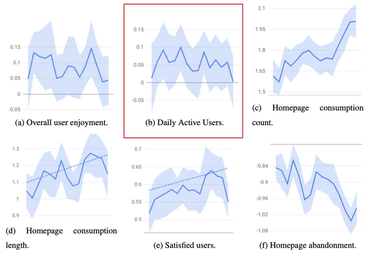


→ What if we structure the prediction around a ***higher-order*** space, i.e., **intents** [Fishbein and Ajzen 1977]?

Missed opportunity: Incorporating an **intent-based structure** into black-box Recsys

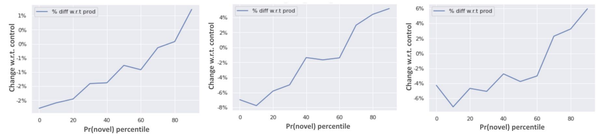
Results

We implemented the framework with the following intents: **Novelty intent** and **familiarity intent**. We observed significant movements in topline business metrics.



	Metric name	Relative change w.r.t. control
Overall metrics	Overall user enjoyment	+0.09%**
	Daily active users (DAU)	+0.05%***
	Homepage consumption count	+1.77%***
	Homepage consumption length	+1.15%***
	Satisfied users	+0.58%***
	Homepage abandonment	-0.92%***
Novelty metrics	Novel creator consumption	+0.72%***
	Repeated novel consumption	+0.10%***
	Unique user-cluster pairs	+0.36%***
Familiarity metrics	Familiar item consumption	+15.92%***
	Familiar creator consumption	+2.77%***

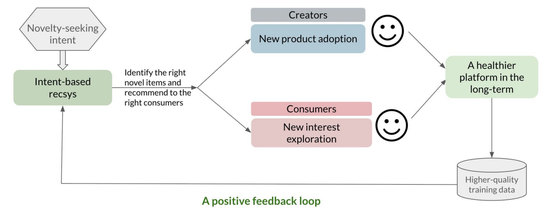
A better intent-aware recommendation experience:



(a) Novel creator impressions. (b) Novel creator consumption. (c) Novel creator CTR.

Higher *P(novelty intent)* → More novel videos recommended → Higher CTR on these videos

Improved Long-Term Outcomes



Conclusion

Our work demonstrates that even when the full data-generating process (DGP) is unknown, identifying some characteristics of it can help the model generalize. Specifically, contrary to the common belief that structure limits model flexibility,imposing a structure that is aligned with the DGP can, in fact, improve the performance of black-box ML systems.

Contact

Yuyan Wang: yuyanw@stanford.edu
Assistant Professor of Marketing, Kevin J. O'Donohue Family Faculty Scholar for 2024-2025

Stanford Graduate School of Business
655 Knight Way
Stanford, CA 94305
United States

References

1. Agrawal, Rakesh, et al. "Diversifying search results." Proceedings of the second ACM international conference on web search and data mining. 2009.
2. Wang, Yuyan, et al. "Surrogate for long-term user experience in recommender systems." Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining. 2022
3. Wang, Yuyan, et al. "Beyond Item Dissimilarities: Diversifying by Intent in Recommender Systems." Proceedings of the 31th ACM SIGKDD conference on knowledge discovery and data mining (to appear).