A Bayesian-DLM-CF Framework for Real-Time Display Advertising

Michael Els[1] David Banks[2]

[1]InMarket mels@inmarket.com

[2]Duke University banks@stat.duke.edu

AdKDD Workshop @ SIGKDD 2025 · 4 Aug 2025

Problem Setup

How do we predict CTR (or other top funnel KPIs) in programmatic advertising?

Problem Setup

How do we predict CTR (or other top funnel KPIs) in programmatic advertising?

Problems:

- **CTR inertia**: cumulative averages hide sudden site drops.
- **Long-tail noise**: low-impression pairs overweighted in traditional RecSys.
- **Cold start**: new campaigns need guidance before data accrues.
- **Behaviour drift**: user & placement patterns change, some slowly and some suddenly.

Overview of the Proposed Method

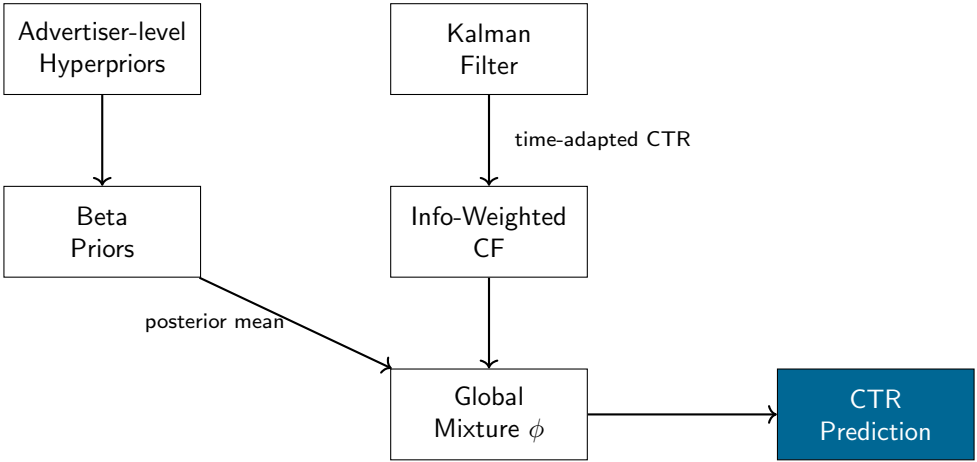
We propose a unified framework integrating:

1. **Beta-based Priors** for sparse or cold-start campaigns
2. **Dynamic Linear Models (DLMs)** for temporal dynamics
3. **Collaborative Filtering (CF)** to exploit item–item or campaign–campaign similarities
4. **Hierarchical Bayesian** sharing across related campaigns

Goal:

- Achieve robust, adaptive CTR estimates that avoid *over-serving* bad items.
- Provide a strong baseline for optional reinforcement learning layers.

High-Level Pipeline



Beta Priors — Uncertainty-Aware Smoothing

$$\text{CTR}_{u,i} \sim \text{Beta}(\alpha_{u,i}, \beta_{u,i})$$

$$\alpha_{u,i} = c_{u,i} + w_\alpha \alpha_0$$

$$\beta_{u,i} = n_{u,i} - c_{u,i} + w_\beta \beta_0$$

- Posterior mean handles sparse counts.
- w_α, w_β pooled hierarchically to help cold-starts.

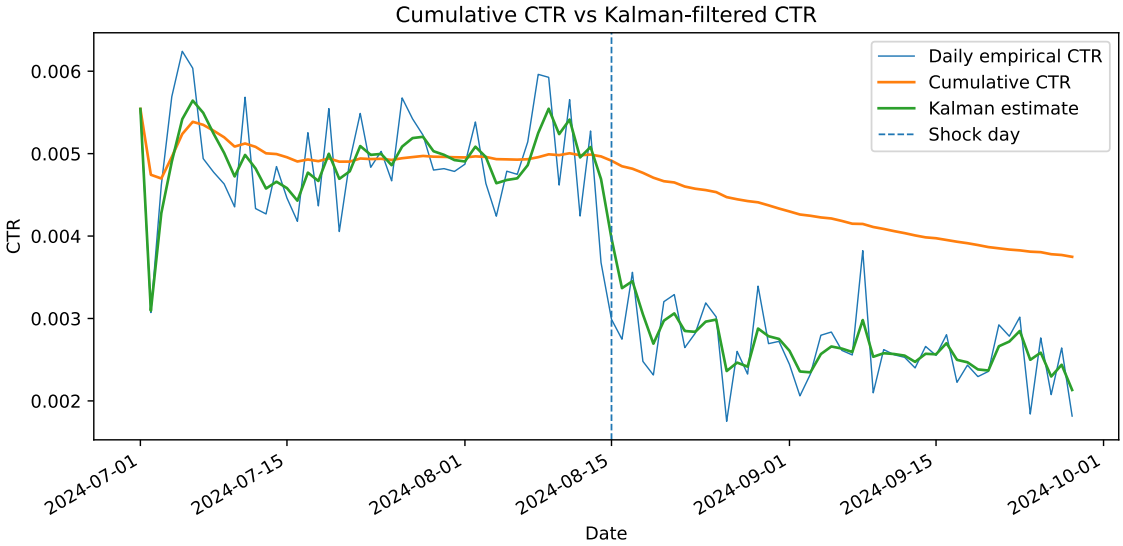
Kalman Filter — State-Space Model

$$\underbrace{x_t}_{\text{latent CTR}} = x_{t-1} + \omega_t, \quad \omega_t \sim \mathcal{N}\Big(0, Q\Big)$$

$$\underbrace{z_t}_{\text{emp. CTR}} = x_t + \nu_t, \quad \nu_t \sim \mathcal{N}\Big(0, R_t\Big)$$

- **Predict**: $\hat{x}_{t|t-1} = x_{t-1}, P_{t|t-1} = P_{t-1} + Q$
- **Kalman Gain**: $K_t = \dfrac{P_{t|t-1}}{P_{t|t-1} + R_t}$
- **Update**: $\hat{x}_t = \hat{x}_{t|t-1} + K_t(z_t - \hat{x}_{t|t-1})$

Kalman Beats Cumulative CTR After a Shock



Vanilla Item-Based CF

$$\mathsf{sim}(i,j) = \frac{\sum_u (x_{u,i} - \bar{x}_i)(x_{u,j} - \bar{x}_j)}{\sqrt{\sum_u (x_{u,i} - \bar{x}_i)^2}\,\sqrt{\sum_u (x_{u,j} - \bar{x}_j)^2}}$$

$$\hat{x}_{u,i} = \frac{\sum_{j \in \mathcal{N}(i)} \mathsf{sim}(i,j)\, x_{u,j}}{\sum_{j \in \mathcal{N}(i)} |\mathsf{sim}(i,j)|}$$

- Treats every pair equally $\rightarrow$ long-tail noise.

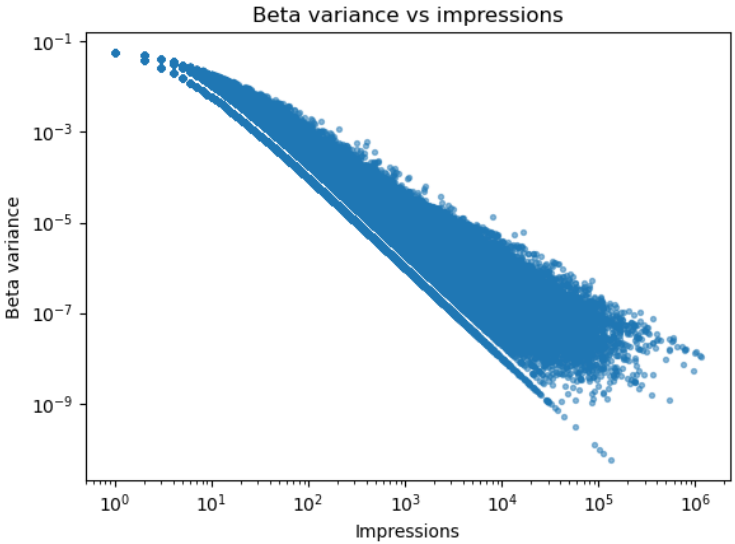
Information-Weighted CF (Ours)

$$w_{u,i} = \frac{(\alpha+\beta)^2(\alpha+\beta+1)}{\alpha\beta} = \frac{1}{\mathsf{Var}[\mathsf{Beta}(\alpha,\beta)]}$$

$$\mathsf{sim}_w(i,j) = \frac{\sum_u w_{u,i} w_{u,j} (x'_{u,i} x'_{u,j})}{\sqrt{\sum_u (w_{u,i} x'_{u,i})^2}\,\sqrt{\sum_u (w_{u,j} x'_{u,j})^2}}$$

- High-variance pairs (few impressions) get tiny weights $\rightarrow$ clean neighbours.
- Top-k truncation keeps inference linear in k.

Beta Variances



Global Mixture Solves Cold Start

$$x_{\text{final}} = \phi \, \frac{\alpha}{\alpha+\beta} + (1-\phi)\, \widetilde{x}_{\text{CF}}$$

- One scalar ϕ learned nightly $\rightarrow$ tiny footprint.
- Cold-start campaigns borrow strength from advertiser hyperpriors.

90-Day DSP Benchmark (Test-Set)

Method	Log-Loss	MSE
Beta+DLM+CF	**0.025234**	**0.000144**
Beta+CF	0.029487	0.000185
DLM+CF	0.030523	0.000523
Vanilla CF	0.035877	0.000654
Random Forest	0.042545	0.000204
XGBoost	0.043481	0.000204

Why One Big Bayesian Model Won't Work

- **No full conjugacy**: Beta–Binomial and Gaussian pieces are conjugate locally, but CF injects $\mathcal{O}(N^2)$ item–item dependencies + global $\phi \rightarrow$ joint posterior has no closed form.
- **Non-conjugate Bayes infeasible**: MCMC/VI would need to sample or optimize $|\mathcal{U}| \times |\mathcal{I}| \times T$ latent states *plus* similarity weights; memory and time blow up (billions of parameters).
- **Modular win**: Keep conjugacy where cheap (Beta, Kalman), treat CF + mixture as deterministic transforms $\rightarrow$ retains interpretability real-time latency

Key Takeaways

- Beta priors tame sparsity; Kalman tames drift.
- Inverse-variance CF exploits cross-campaign structure while accounting for item information.
- Global shortcuts and Bayesian conjugacy keep computation efficient.