SMARTINFER
Case Study

Quantifying the Demand–Supply Gap in E-Commerce Using Topic Models

How Walmart used information-theoretic methods to turn search logs into actionable assortment and demand generation strategy

Author	Published	Co-Authors	Context
Anjan Goswami	WWW 2019, San Francisco	P. Mohapatra, C. Zhai	Walmart E-Commerce

The Problem

Demand generation and assortment selection are two of the largest investment areas in retail. Traditionally, both rely on historical sales data and merchandiser intuition. But sales data only captures transactions that actually happened—it is silent about demand that was never fulfilled because the right product was never available, and about inventory that sits unsold because no customer ever looked for it.

At Walmart, with millions of products and hundreds of millions of search queries, the volume and variety of customer intent signals were well beyond meaningful human interpretation. The marketplace faced a specific set of compounding problems:

- **Queries without matching products:** Customers searched for items the catalog did not carry, representing unmet demand invisible to sales-based planning.
- **Products without matching queries:** A large segment of the catalog had near-zero impressions—items the search system never surfaced because no customer query aligned with them.
- **No scalable framework:** Individual category managers and internal seller units each operated with local visibility. No system connected what customers wanted (demand) with what the marketplace offered (supply) at the topic level.

The question was not whether a gap existed—it was how to systematically discover, quantify, and act on it across the entire marketplace.

The Approach

Modeling Demand and Supply as Topic Distributions

We treated the problem as an information-theoretic one. Customer search queries represent a distribution over latent topics—the demand signal. The product catalog, represented through item titles and descriptions, represents a separate distribution over those same latent topics—the supply signal. The gap between the two distributions is the marketplace inefficiency.

We constructed separate topic models using Latent Dirichlet Allocation (LDA) on two corpora: a representative sample of historical search queries from Walmart.com, and the digitized product catalog. Each corpus produced a distribution over discovered topics, enabling direct comparison.

Quantifying the Gap with KL Divergence

To move from visualization to measurement, we defined a formal gap metric based on **Kullback–Leibler divergence** between the query topic distribution (demand) and the product topic distribution (supply). The KL divergence decomposes per topic, producing a signed, ranked list:

Gap Signal	Interpretation	Action
D(topic) >> S(topic)	Excess demand: customers search for this topic far more than the	Route to assortment selection and buyers for catalog expansion

	catalog serves it	
S(topic) >> D(topic)	Excess supply: inventory exists in this topic space but customers rarely search for it	Route keywords to demand generation team; evaluate for delisting
D(topic) ≈ S(topic)	Equilibrium: supply matches demand in this topic	Optimize ranking and conversion within existing assortment

The per-topic decomposition made the gap actionable: each topic came with its constituent keywords, which provided the exact vocabulary for demand generation campaigns or buyer search criteria.

Closing the Loop: From Insight to Action

The gap analysis fed into three operational pipelines, each targeting a different stakeholder within Walmart's marketplace organization:

Assortment Selection. Topic clusters with excess demand were mapped to specific product categories and routed to buyers. Because Walmart's internal business units each operate semi-independently—selecting and submitting their own inventory—the gap analysis gave buyers a data-driven view of unmet customer demand that was previously invisible to them. The extracted keywords served as search criteria for sourcing from suppliers and competitor benchmarking.

Demand Generation. Topic clusters with excess supply—products in the catalog that matched no active query cluster—were routed to the marketing and demand generation teams. The topic model provided not just the product categories to promote, but the specific keyword vocabulary to use in advertising, SEO, and campaign targeting. This transformed demand generation from an intuition-driven exercise into a data-informed one.

Search and Discovery. A subset of items with low impressions had topics that overlapped with active query clusters but were not being retrieved. This identified retrieval and indexing failures—products the system should have surfaced but didn't—enabling targeted fixes to the search pipeline independent of the assortment or marketing interventions.

Results and Impact

The supply–demand gap framework produced measurable outcomes across all three action pipelines:

- **The gap metric correlated with revenue,** validating that topics with high KL divergence corresponded to genuine business opportunity. This provided a principled way to prioritize where to invest in assortment expansion versus demand generation.
- **Buyer teams received actionable gap reports** with specific product categories and keywords representing unmet demand, replacing anecdotal sourcing decisions with quantified opportunity sizing.
- **Demand generation campaigns were targeted using the exact keyword vocabulary** extracted from excess-supply topics, improving the precision of marketing spend.
- **The framework contributed to the broader search and marketplace optimization effort** that delivered +23% revenue uplift and +17% conversion improvement across Walmart e-commerce search.

Technical Contribution

The work was published at The Web Conference 2019:

> Goswami, A., Mohapatra, P., and Zhai, C. (2019). *"Quantifying and Visualizing the Demand and Supply Gap from E-commerce Search Data using Topic Models."* Companion Proceedings of The Web Conference (WWW '19), pp. 348–353. ACM.

The key methodological contribution was demonstrating that topic models on search logs and product catalogs, combined with KL divergence as a gap metric, provide a scalable, unsupervised framework for supply–demand analysis that does not require sales transaction data. This means the framework can identify gaps before they manifest as lost revenue—a forward-looking capability that sales-based methods cannot provide.

Broader Applicability

The demand–supply gap framework is not specific to Walmart or to retail. Any marketplace with a query-based discovery mechanism and a catalog of offerings can apply the same methodology: job marketplaces (what candidates search for versus what employers post), content platforms (what users search for versus what creators produce), B2B procurement (what buyers need versus what suppliers list). The core insight—that search behavior reveals demand independently of transaction history—generalizes to any two-sided market where discovery happens through search.