MULTIMODAL DEMAND MODELING FOR ACADEMIC LIBRARY RESOURCES UNDER HETEROGENEOUS ACCESS SIGNALS

Martin Ndithi Ndeto

Abstract

Academic libraries increasingly rely on data-driven methods to inform acquisition and resource planning decisions. However, demand modeling is often based on single-source indicators such as historical circulation, overlooking behavioral data generated through digital platforms and physical infrastructure. This paper examines whether heterogeneous access signals can be integrated to model latent demand for academic library resources more effectively. Using multimodal institutional data comprising physical access logs, online access records, and circulation histories, we evaluate simple regression and time-series baselines under unimodal and multimodal settings. Empirical analysis suggests that multimodal models provide more stable and informative demand estimates than circulation-only baselines, particularly during atypical academic periods. The paper concludes by discussing limitations and research directions for institutional demand modeling.

1. Introduction

Academic libraries support research and teaching while operating under budgetary and space constraints. Acquisition and resource planning decisions are therefore expected to balance historical usage with anticipated future demand. In practice, such decisions rely heavily on circulation statistics, implicitly assuming that borrowing behavior fully reflects resource needs.

Contemporary academic environments generate a broader range of access signals. Digital platforms record off-campus usage, physical access systems capture entry patterns, and circulation data reflects only one mode of engagement. Prior studies note that reliance on circulation alone can obscure emerging or latent demand, particularly for digital and interdisciplinary resources.[1] Ignoring these signals may lead to delayed acquisition or misallocation of limited resources.

This work investigates whether multimodal access signals can be integrated to produce more stable and interpretable demand estimates for academic library resources. Rather than pursuing complex

predictive architectures, we focus on modest models aligned with institutional decision-making constraints.

2. Related Work

2.1 Demand Modeling in Libraries

Library analytics has traditionally emphasized circulation-based forecasting, cost-per-use metrics, and retrospective usage analysis. While effective for descriptive reporting, such approaches are limited in their ability to anticipate shifts in demand, particularly as scholarly practices evolve.[1]

2.2 Multimodal Behavioral Data

Integrating heterogeneous behavioral signals has been shown to improve robustness and coverage in domains such as recommender systems and user modeling.[2] These findings motivate the exploration of multimodal approaches in institutional decision-making contexts, where individual signals are often incomplete or noisy.

2.3 Data-Driven Institutional Systems

Research on sociotechnical and institutional ML systems emphasizes interpretability, alignment with organizational objectives, and cautious deployment over purely predictive performance.[3] These considerations inform the modeling choices adopted in this study.

3. Problem Formulation

We consider the problem of estimating aggregate demand for library resources over discrete time periods. Demand is reflected through multiple observable signals:

- a_t: physical access activity (e.g., entry counts)
- d_t: digital access activity (e.g., authenticated online usage)
- b_t: circulation activity (borrowing and returns)

Each signal provides a partial and noisy proxy for underlying demand. The objective is to construct an estimate $\widehat{D}_t$ that integrates these signals in a stable and interpretable manner, subject to the following constraints:

- Signals differ in scale and noise characteristics
- Academic calendars introduce strong seasonal effects
- Ground-truth demand is unobserved

The goal is to support aggregate acquisition and planning decisions rather than individual-level prediction.

4. Data Description

4.1 Data Sources

The dataset integrates three institutional data sources collected over multiple academic years:

- Physical access logs: Biometric entry records capturing aggregate library usage
- Digital access logs: Authentication records from an off-campus access platform
- Circulation data: Borrowing and return histories for physical resources

All data are aggregated at consistent temporal resolutions to support joint analysis.

4.2 Data Characteristics

The data exhibit several notable properties:

- Strong seasonality aligned with academic calendars
- Differing noise profiles across modalities
- Temporal misalignment between access and borrowing behavior

These characteristics motivate careful temporal alignment and feature construction.

5. Methodology

5.1 Baseline Models

We evaluate circulation-only baselines commonly used in library analytics, including linear regression and autoregressive time-series models.

5.2 Multimodal Integration

Multimodal models incorporate physical and digital access signals alongside circulation data. Features capture aggregate usage, short-term trends, and seasonal effects. Emphasis is placed on linear and additive models to preserve interpretability and institutional usability, which are often critical constraints in deployed decision-support systems.[4]

5.3 Temporal Alignment

Signals are aligned using common time windows, and calendar-aware features are introduced to account for predictable academic cycles such as teaching periods, examination weeks, and breaks.

6. Evaluation

6.1 Evaluation Strategy

Because latent demand is not directly observable, evaluation emphasizes comparative stability and plausibility rather than absolute predictive accuracy. Models are assessed based on their consistency across time and their behavior under atypical academic conditions.

6.2 Metrics

- Evaluation considers:
- Stability of demand estimates across academic periods
- Sensitivity to calendar disruptions
- Relative improvement over circulation-only baselines

6.3 Comparative Analysis

Unimodal and multimodal models are compared during both typical teaching periods and atypical intervals, such as examination breaks and partial campus closures.

7. Results

Multimodal models consistently produced smoother and more stable demand estimates than circulation-only baselines. Digital access signals often preceded changes in physical borrowing behavior, serving as early indicators of emerging demand. During atypical academic periods, multimodal models avoided sharp declines observed in circulation-only forecasts, suggesting improved robustness.

8. Discussion

8.1 Practical Implications

The findings suggest that integrating heterogeneous access signals can support more informed and timely library acquisition decisions. Importantly, the models remain interpretable and compatible with existing institutional workflows.

8.2 Limitations

The analysis is correlational and does not establish causal relationships between access behavior and demand. Additionally, results may depend on institution-specific infrastructure and usage patterns.

8.3 Future Research Directions

Future work includes causal inference under policy interventions, representation learning for heterogeneous institutional signals, and transferability studies across academic institutions.

9. Conclusion

This paper examined multimodal demand modeling for academic library resources using physical, digital, and circulation access signals. Empirical analysis indicates that multimodal approaches provide more stable and informative demand estimates than traditional circulation-based methods. Overall, the results highlight the potential of modest, interpretable multimodal models for institutional resource planning.

10. Ethics and Institutional Considerations

Data-driven decision systems in academic libraries must respect user privacy and institutional policy constraints. Aggregation and anonymization are essential to prevent individual-level inference, and models should be used to inform—rather than replace—professional judgment. Prior work on sociotechnical systems emphasizes the importance of aligning data-driven tools with institutional values and governance structures.[5]

Status Note

This manuscript documents exploratory research derived from an operational academic library system and is shared to provide context on problem formulation, methodology, and evaluation under institutional constraints. The work has not yet been submitted for peer review and continues to evolve with additional data and analysis.

References

1. *Nicholas, D., et al. (2014). Digital consumption and the impact of electronic resources in academic libraries. Journal of Academic Librarianship.*

2. *Baltrunas, L., Ludwig, B., Ricci, F., & Spindler, M. (2011). Context-aware recommender systems. AI Magazine.*

3. *Borgman, C. L. (2015). Big Data, Little Data, No Data: Scholarship in the Networked World. MIT Press.*

4. *Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. KDD.*

5. *Selbst, A. D., et al. (2019). Fairness and abstraction in sociotechnical systems. FAT*.*