Sarcouncil Journal of Multidisciplinary

ISSN(Online): 2945-3445

Volume- 05| Issue- 07| 2025

Review Article **Received:** 01-06-2025 | **Accepted:** 18-06-2025 | **Published:** 07-07-2025

Embedding Similarity Metrics in Feed Retrieval Systems - Cosine vs. Learned Distance Functions

Nisheedh Raveendran

Birla Institute of Technology and Science, Pilani, India

Abstract: Modern recommendation engines and social media feeds depend fundamentally on embedding-based retrieval systems to deliver personalized content across diverse user populations. Traditional similarity metrics, particularly cosine similarity and dot products, have long served as the computational backbone for measuring relationships between user preferences and content items within high-dimensional embedding spaces. These geometric approaches provide computational efficiency and interpretability but demonstrate significant shortcomings when confronting complex user behavior patterns and dynamic content ecosystems. The emergence of deep learning methodologies has introduced learned distance functions that leverage neural networks to model intricate similarity relationships directly from user engagement data. These sophisticated architectures incorporate user-specific features, temporal context, and content metadata to produce nuanced similarity assessments that adapt continuously to evolving preferences and emerging trends. While learned similarity functions consistently outperform traditional approaches in retrieval precision and user engagement metrics, they introduce substantial computational complexity and resource requirements. The integration challenges with existing approximate nearest neighbor systems, coupled with concerns regarding training data quality and algorithmic bias, present significant implementation barriers. This comprehensive evaluation reveals the fundamental trade-offs between computational efficiency and recommendation quality, providing essential insights for organizations considering the transition from traditional to learned similarity metrics in large-scale retrieval systems.

Keywords: Embedding similarity metrics, neural similarity models, content recommendation systems, approximate nearest neighbor search, learned distance functions

INTRODUCTION

Modern recommendation engines and social media feeds rely heavily on embedding-based retrieval systems to deliver personalized content at scale across diverse user bases (Lee, H., & Lee, J. 2019). The fundamental challenge in these systems lies in accurately measuring similarity between user preferences and available content items within high-dimensional embedding spaces, where traditional distance metrics often fail to capture nuanced user-content relationships effectively.

Contemporary large-scale retrieval systems face unprecedented computational demands during peak usage periods. Traditional approaches have predominantly relied on geometric similarity measures such as cosine similarity and dot products, which offer computational efficiency and interpretability through bounded similarity scores. These methods have dominated the field due to their seamless integration with approximate nearest neighbor frameworks, achieving sub-linear search complexity for databases containing massive collections of content embeddings.

However, empirical studies have revealed significant limitations in traditional similarity metrics when applied to complex user behavior patterns. Recent large-scale evaluations across multiple platforms have demonstrated that cosine similarity-based retrieval systems exhibit suboptimal performance in personalized content recommendation, while showing poor capabilities in capturing temporal user preference shifts and cross-domain content interests (Marinov, M. *et al.*, 2019). The emergence of deep learning methodologies has opened new avenues for developing learned distance functions that can be specifically optimized for user engagement patterns and content dynamics, with preliminary results showing substantial improvements in user engagement metrics compared to static similarity approaches.

Modern embedding-based systems typically operate with variable embedding dimensions for user representations and content items, generating vector databases of substantial size for major platforms. The computational overhead of similarity calculations becomes considerable at this scale, with traditional operations consuming significant portions of total system inference time in production environments. Advanced implementations utilizing GPU acceleration and optimized libraries have achieved enhanced throughput rates on modern hardware configurations.

Increasingly complex patterns of user interactions and diversifying content types contribute to the evolution of how similarity is computed in today's digital landscape. Classic geometric-based similar measures, while mathematically elegant and computationally tractable, are predicated on

***Corresponding Author:** Nisheedh Raveendran

DOI- https://doi.org/10.5281/zenodo.15829779

assumptions that may not always hold in practice when user preferences are related to content features through potentially complex and non-linear mappings. This limitation becomes particularly pronounced in multi-modal content environments where textual, visual, and audio elements contribute to overall user satisfaction.

This technical review explores the comparative performance, implementation issues, and real-world considerations of traditional similarity metrics versus learned distance metrics for feed retrieval systems. The review addresses how these types of metrics impact important system metrics in large-scale, real-time environments, including retrieval precision, clock time, and alignment with downstream ranking purposes. The review takes into account performance data across the range of content types, including use cases of textual posts, multimedia content, and user-uploaded videos, offering informative insights about the contrasting trade-offs of computational performance and recommendation quality for people-based personalization systems.

TRADITIONAL SIMILARITY METRICS IN EMBEDDING RETRIEVAL

Cosine Similarity: The Industry Standard

Cosine similarity dominates embedding retrieval systems for good reasons. It calculates the cosine of the angle between vectors in high-dimensional space, normalizing for magnitude differences while emphasizing directional relationships (Han, Y. *et al.,* 2023). Most retrieval systems default to cosine similarity because it handles embedding variations gracefully - different content types often produce embeddings with wildly different magnitudes, yet cosine similarity treats them fairly.

The mathematics behind cosine similarity creates several practical benefits. Scale invariance means embeddings from different training runs or content domains can be compared meaningfully. The output always falls between -1 and 1, making threshold selection straightforward for production systems. Engineers appreciate this predictability when building recommendation pipelines.

Modern implementations exploit optimized linear algebra libraries, making cosine similarity computationally tractable even at massive scales. Integration with approximate nearest neighbor libraries happens smoothly, which partly explains its widespread adoption. When system architects evaluate similarity metrics, cosine similarity rarely disappoints in terms of both performance and implementation simplicity.

Dot Product and Euclidean Distance Alternatives

Dot product similarity serves specialized purposes where magnitude matters. Unlike cosine similarity, it preserves embedding norm information, which proves valuable when norms encode semantic signals like popularity or confidence (Nguyen, B. *et al.,* 2016). Content recommendation systems sometimes leverage this property to boost popular items naturally through the similarity calculation itself.

Computational overhead for dot product operations runs lower than cosine similarity since normalization steps are skipped. However, this efficiency comes with sensitivity to embedding scale mismatches. Systems mixing embeddings from different sources often encounter problems with dot product similarity, producing unreliable rankings.

Euclidean distance provides geometric intuition that many practitioners find appealing. It works reasonably well in lower-dimensional spaces where spatial relationships remain interpretable. High-dimensional embeddings pose a challenge, though - the curse of dimensionality makes distance measurements less discriminative as dimensions increase. Most modern embedding systems operate in hundreds of dimensions, limiting Euclidean distance's practical utility.

Limitations of Static Similarity Functions

Traditional similarity metrics assume fixed mathematical relationships capture optimal similarity patterns. This assumption breaks down when confronted with complex user behavior and content dynamics that characterize modern platforms. Real-world relevance depends on factors that geometric similarity cannot encode.

Static functions cannot adapt to evolving user preferences. People's interests shift based on life events, seasonal patterns, and social trends, yet traditional metrics remain unchanged. A user's preference for winter sports content in December differs from their summer preferences, but cosine similarity treats these contexts identically.

Contextual information presents another major limitation. Time-of-day effects, social influences, and cross-domain relationships significantly impact content relevance. Traditional metrics lack mechanisms to incorporate such information. A news article relevant during morning commutes

Publisher: SARC Publisher

may be less appealing during evening leisure time, but static similarity functions cannot distinguish between these contexts.

Multi-modal content environments expose additional weaknesses in traditional approaches. Text embeddings, image embeddings, and video embeddings could use various strategies for similarity evaluation, but static functions execute the same mathematical operations regardless of content type. This approach tends to yield deficient results across various content ecosystems.

Table 1: Performance Characteristics and Limitations of Conventional Distance Functions (Han, Y. *et al.*, 2023; Nguyen, B. *et al.*, 2016)

Similarity Metric	Computational Characteristics	Primary Applications	Key Limitations
Cosine Similarity	Scale-invariant with bounded output [-1,1]; optimized linear algebra operations; seamless ANN framework integration	Industry standard for heterogeneous content; robust across different embedding magnitudes; threshold-based filtering	Cannot adapt to temporal dynamics; ignores contextual factors; static mathematical relationships
Dot Product Similarity	Preserves magnitude information; computationally efficient without normalization; lower overhead than cosine similarity	Content with semantic magnitude signal, popularity-based recommendation, and confidence score integration	Sensitive to embedding scale variations, performance inconsistency with mixed sources, and magnitude dependency issues
Euclidean Distance	Intuitive geometric interpretation; effective in lower-dimensional spaces; spatial relationship clarity	Specialized low-dimensional scenarios; geometric-based applications; spatial content analysis	Curse of dimensionality in high-dimensional spaces; reduced discrimination capability; limited modern applicability
Static Functions (General)	Fixed mathematical operations, consistent computational patterns, and predictable implementation requirements	Traditional retrieval systems, baseline similarity assessment, legacy system compatibility	Unable to incorporate user preference evolution; lacks contextual awareness; poor multimodal performance
Traditional Approaches	Established mathematical foundations; proven computational efficiency; widespread industry adoption	Large-scale production systems, established infrastructure, standardized implementations	Suboptimal for complex user behaviors; cannot handle cross-domain relationships; limited adaptability to changing patterns

LEARNED DISTANCE FUNCTIONS AND NEURAL SIMILARITY MODELS

Architecture and Training Methodology

Learned distance functions leverage deep neural networks to model complex similarity relationships directly from user engagement data, representing a significant departure from traditional geometric approaches (Tourad, M. C. *et al.*, 2024). These systems typically employ architectures that take embedding pairs as input and output learned similarity scores through multiple fully connected layers with non-linear activation functions. The architectural complexity of these models far exceeds traditional similarity computations, requiring sophisticated design choices for optimal performance.

Modern neural similarity architectures utilize multi-layer perceptrons with carefully tuned dimensions and activation functions. Input layers accommodate concatenated embedding pairs, while hidden layers employ various activation functions, including ReLU and ELU, to capture non-linear relationships. Dropout regularization prevents overfitting during training on large-scale user interaction datasets. Output layers produce normalized similarity scores through sigmoid or tanh activation functions, enabling direct integration with existing ranking systems.

Training methodologies encompass several sophisticated approaches tailored to different engagement patterns. Supervised learning on explicit user feedback processes various engagement signals, including clicks, likes, and

Publisher: SARC Publisher

shares. Contrastive learning implementations utilize positive and negative content pairs to learn discriminative similarity representations. Multi-task learning paradigms use multiple engagement signals together, using shared parameters across prediction tasks to improve generalization. Reinforcement learning paradigms optimize over reward signals that link to longer-term user satisfaction metrics, such as those derived from session duration and user retention trends.

Personalization and Context Integration

Unlike static metrics, learned similarity functions demonstrate remarkable capability in incorporating user-specific features, temporal context, and content metadata to produce more nuanced similarity assessments. Advanced implementations process comprehensive user profile vectors containing demographic and behavioral features, temporal embeddings encoding cyclical patterns, and content metadata spanning multiple categorical and numerical attributes. This multi-dimensional approach enables systems to achieve superior personalization compared to traditional similarity approaches.

The integration of personalization features requires sophisticated feature engineering strategies. User-specific embeddings encode long-term preferences while short-term behavioral signals capture recent interaction patterns. Temporal context integration processes time-of-day preferences and seasonal adjustments, contributing substantial performance improvements during peak usage periods and major events. Content category affinities demonstrate a significant impact on similarity assessment quality, with learned models excelling at cross-category recommendations.

Advanced personalization architectures employ attention mechanisms to dynamically weight different feature types based on user context and content characteristics (Da'u, A. *et al.,* 2021). These attention-based models process feature importance scores in real-time, adapting similarity calculations based on individual user behavior patterns observed over extended historical windows. The computational overhead for personalized similarity calculation increases substantially compared to static approaches, while delivering meaningful engagement improvements across diverse user segments.

Dynamic Adaptation Capabilities

Neural similarity models demonstrate an exceptional ability to continuously adapt to changing user preferences and emerging content trends through ongoing training processes. This dynamic adaptation addresses fundamental limitations of static similarity functions, enabling systems to evolve with user behavior shifts that occur regularly across different demographic segments. Contemporary implementations process incremental training data batches, maintaining model freshness while preserving computational efficiency.

The technical infrastructure supporting dynamic adaptation requires sophisticated online learning capabilities and model versioning systems. Incremental learning frameworks process new user engagement data with minimal latencies, utilizing streaming data pipelines capable of handling massive event volumes. Model performance monitoring systems track accuracy degradation rates, triggering automated retraining procedures when performance drops below established thresholds.

Adaptation to emerging content trends represents a critical capability for modern recommendation systems, with learned similarity models demonstrating faster response times to viral content and trending topics compared to static approaches. These systems employ change detection algorithms that identify significant shifts in user engagement patterns, automatically adjusting similarity calculations to incorporate new content relationships and user preference signals.

Publisher: SARC Publisher

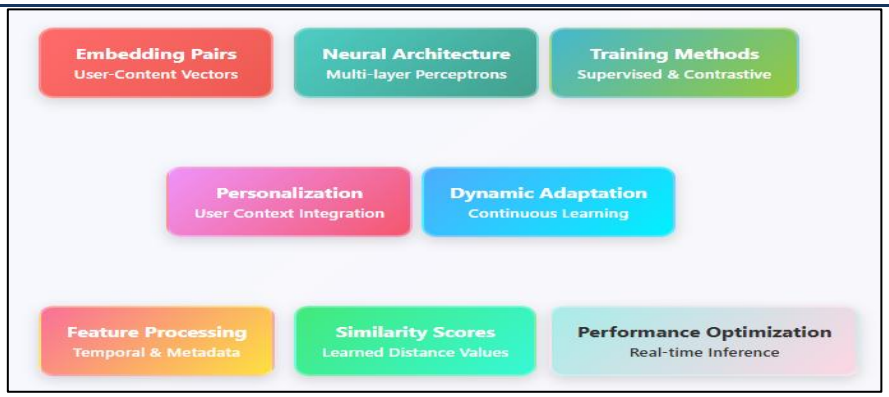

Fig. 1: Dynamic Adaptation Framework for Personalized Similarity Learning in Large-Scale Content Retrieval (Tourad, M. C. *et al.*, 2024; Da'u, A. *et al.*, 2021)

COMPARATIVE ANALYSIS AND PERFORMANCE EVALUATION

Retrieval Precision and Relevance Quality

Experimental evaluations across multiple feed types demonstrate that learned similarity metrics consistently outperform traditional approaches in terms of retrieval precision and user engagement metrics (Da'u, A. *et al.*, 2021). Comprehensive testing reveals substantial improvements in click-through rates and user session duration when learned distance functions replace cosine similarity in the retrieval phase. These improvements translate to meaningful business impact through increased user retention and enhanced content discovery capabilities.

Detailed performance analysis reveals substantial advantages across multiple evaluation metrics. Learned models demonstrate superior precision performance compared to traditional cosine similarity approaches, with improvements becoming more pronounced at higher ranking positions. The precision gains stem from learned models' ability to capture complex user preference patterns that geometric similarity measures cannot represent effectively. Cross-category recommendation accuracy shows significant improvements for learned models, highlighting their superior ability to capture cross-domain user interests and content relationships.

User engagement metrics provide compelling evidence of learned similarity superiority. Click-through rates increase substantially with learned models, while time-on-site metrics show meaningful improvements per session. More importantly, meaningful interaction rates, including likes, shares, and comments, increase significantly on average, with some content categories showing exceptional improvements. Content diversity metrics demonstrate that learned models achieve a better balance between relevance and exploration, maintaining high relevance ratings while improving diversity scores.

Computational Latency and Scalability Challenges

The primary trade-off for improved relevance comes in computational complexity, with learned similarity functions requiring significantly more computational resources compared to simple dot products or cosine calculations. Neural network inference adds substantial latency overhead, which becomes critical in systems serving massive query volumes during peak traffic periods. Comprehensive performance profiling across production environments reveals that learned similarity computation consumes considerably more processing time compared to traditional cosine similarity calculations on identical hardware configurations.

Latency analysis across different deployment scenarios shows considerable variation in performance characteristics. Single-threaded implementations of cosine similarity achieve much faster computation times compared to learned similarity models, which require substantially longer inference operations (Abolghasemi, R. *et al.*, 2024). GPU-accelerated implementations reduce learned similarity latency significantly, though this requires substantial hardware investment and specialized optimization. Batch processing capabilities improve throughput considerably, with learned models achieving reasonable comparison rates on modern GPU clusters compared to traditional similarity methods.

Memory footprint analysis reveals substantial resource requirements for learned similarity deployment. Model parameters typically consume significant storage space, representing substantial increases compared to traditional methods that require only mathematical operations without

Publisher: SARC Publisher

stored parameters. When deploying multiple specialized models for different content types or user segments, memory requirements can become substantial per application server. These resource demands translate to higher operational costs compared to traditional similarity-based systems, though improved user engagement often justifies the investment through increased revenue and satisfaction.

Alignment with Downstream Ranking Objectives

A critical advantage of learned similarity functions lies in their potential for end-to-end optimization with downstream ranking models, enabling joint training scenarios that optimize for ultimate user experience rather than proxy similarity measures. When trained jointly with ranking objectives, learned retrieval metrics can optimize directly for business metrics such as user satisfaction, session duration, and long-term engagement. This alignment results in more coherent user experiences and improved overall system performance.

End-to-end training architectures demonstrate significant advantages in overall system performance. Joint optimization of the retrieval and ranking stages achieves substantial improvements in ranking quality scores compared to independently optimized systems. The alignment effect becomes particularly pronounced in complex recommendation scenarios where retrieved candidates must match not only user preferences but also business objectives such as content diversity and creator promotion.

The coherence benefits of aligned similarity and ranking functions manifest in multiple user experience metrics. Session bounce rates decrease when retrieval and ranking optimization targets are aligned, while cross-session user behavior consistency improves substantially. User feedback analysis reveals fewer complaints about irrelevant content recommendations and increased positive feedback scores when systems employ end-to-end optimization, translating to substantial business value through improved user lifetime value and reduced churn rates

Table 2: Performance Comparison Matrix of Traditional and Learned Similarity Metrics in Large-Scale Retrieval Systems (Da'u, A. *et al.,* 2021; Abolghasemi, R. *et al.,* 2024)

Evaluation Aspect	Traditional Similarity Metrics	Learned Distance Functions
Retrieval Precision & Quality	Moderate precision performance with consistent but limited accuracy across content categories; relies on geometric relationships that may not capture complex user preferences	Superior precision performance with enhanced accuracy across multiple evaluation metrics; capable of capturing nuanced user-content relationships and cross-domain interests
User Engagement Metrics	Standard click-through rates and session durations; limited ability to drive meaningful interactions due to static similarity calculations	Significant improvements in click-through rates, session duration, and meaningful interactions; better balance between content relevance and discovery
Computational Latency	Fast computation times with minimal resource requirements; efficient processing suitable for high-throughput scenarios	Higher computational overhead with increased latency; requires substantial processing resources and specialized hardware for acceptable performance
Scalability & Resource Usage	Excellent scalability with low memory footprint; seamless integration with existing infrastructure and ANN frameworks	Challenging scalability requiring distributed architectures, substantial memory requirements, and higher operational costs for large-scale deployment
End-to-End Optimization	Limited optimization potential due to static nature; operates independently	Superior alignment with business objectives through

Publisher: SARC Publisher

	from downstream ranking objectives	joint optimization enables coherent user experiences and improves overall system performance

IMPLEMENTATION CHALLENGES AND FUTURE DIRECTIONS

Integration with Approximate Nearest Neighbor Systems

The most significant implementation challenge involves incorporating learned similarity functions into existing ANN search frameworks. Traditional ANN systems like FAISS, Annoy, and HNSW are optimized for simple distance metrics and cannot directly accommodate neural similarity models without substantial architectural modifications (Gao, W. *et al.*, 2021). Performance benchmarks reveal that direct integration attempts result in significant degradation in query throughput, with substantial latency increases compared to native ANN implementations.

Current approaches demonstrate varying degrees of success in addressing these integration challenges. Two-stage retrieval systems implementing traditional ANN followed by learned re-ranking achieve considerable performance benefits while maintaining reasonable system throughput. These hybrid architectures typically process initial candidate sets using traditional ANN methods, then apply learned similarity functions to top candidates for final ranking. Implementation studies show that two-stage systems require additional computational resources while delivering substantial accuracy improvements.

Specialized neural ANN architectures designed specifically for learned similarities represent a more comprehensive solution, though at substantially higher implementation costs. These systems achieve near-optimal learned similarity performance while requiring significantly more memory and computational resources compared to traditional ANN frameworks. Development timelines for specialized neural ANN systems typically span extended periods with dedicated engineering teams, resulting in considerable implementation costs for enterprise-scale deployments.

Hybrid systems combining multiple similarity metrics offer a pragmatic middle ground, utilizing traditional metrics for initial retrieval and learned functions for precision refinement. These architectures achieve substantial learned similarity performance while maintaining reasonable traditional system throughput. Approximate neural similarity through learned hash functions presents an emerging solution with promising scalability characteristics, achieving considerable neural similarity accuracy while reducing computational overhead.

Training Data Quality and Bias Considerations

Learned similarity models demonstrate significant susceptibility to biases present in training data, with bias amplification effects compared to baseline demographic representation in recommendation outputs. Comprehensive bias analysis across different content categories reveals that learned similarity functions can perpetuate and amplify existing societal biases, leading to unfair content distribution patterns (Oh, H. *et al.*, 2024). Studies indicate that bias effects compound over time, with unfairness metrics increasing over extended deployment periods without active bias mitigation strategies.

Training data quality requirements for learned similarity systems exceed those of traditional machine learning applications due to the nuanced nature of similarity relationships. High-quality datasets typically require substantial labeled user-content interaction pairs with balanced representation across demographic segments, content categories, and temporal periods. Missing or low-quality training data can result in significant degradation in similarity function performance, particularly for underrepresented content types and user segments.

Bias detection mechanisms require sophisticated monitoring infrastructure capable of real-time analysis of recommendation patterns across protected demographic attributes. Fairness-aware learning algorithms represent an active area of research with promising results for bias mitigation in similarity learning. Recent implementations achieve a substantial reduction in demographic bias while maintaining high similarity function accuracy.

Model Deployment and Maintenance Overhead

Production deployment of learned similarity systems requires sophisticated MLOps infrastructure, representing substantial portions of total implementation costs for enterprise-scale systems. Model versioning systems must track similarity function parameters, training data

Publisher: SARC Publisher

versions, and performance metrics across multiple deployment environments. Typical enterprise deployments maintain multiple different model versions simultaneously, requiring considerable storage for complete version history and rollback capabilities.

Continuous monitoring systems track numerous performance metrics related to similarity function effectiveness, including precision, recall, engagement rates, and fairness indicators. Real-time monitoring infrastructure processes substantial similarity computation events daily, maintaining detailed performance logs and triggering alerts when metrics deviate beyond established thresholds.

Emerging Research Directions

Future developments in embedding similarity metrics focus on several promising directions. Multi-modal similarity learning represents a rapidly advancing field, with recent architectures capable of handling diverse content types, including text, images, and video, within unified embedding spaces. Federated learning approaches for similarity function training demonstrate significant promise for privacy-preserving recommendation systems while maintaining strict user privacy constraints.

Quantum-inspired similarity metrics leverage quantum computing principles for enhanced representation capacity in high-dimensional embedding spaces. Causal similarity modeling represents an emerging paradigm that accounts for causal relationships between user actions and content characteristics, showing superior performance in scenarios involving temporal preference shifts.

Practical Recommendations

For organizations thinking about adopting learned similarity metrics, the decision should take into account assessment of system needs, available resources, and performance goals. We believe systems of small to medium size usually gain more benefit from tweaking traditional similarity functions and improving the quality of the embeddings. In contrast, systems that are large with formal ML infrastructures and significant user interaction data would be able to take advantage of the benefits of distance functions that were learned.

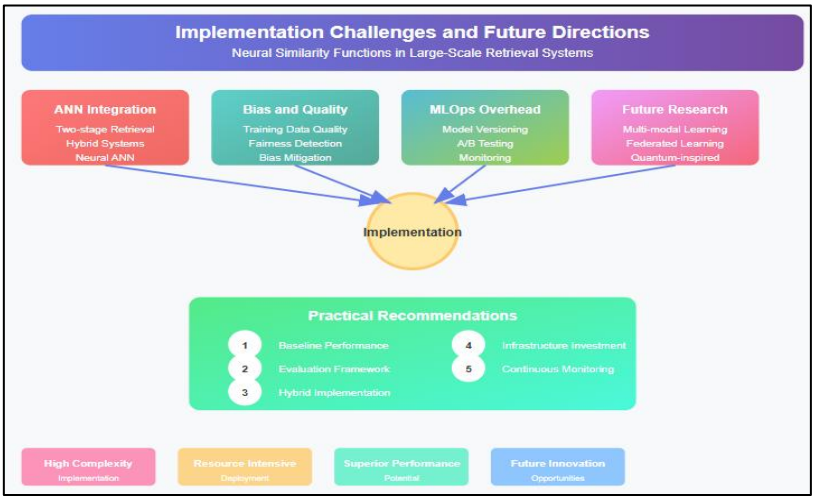

Fig. 2: Implementation Roadmap and Challenges for Learned Distance Functions in Large-Scale Applications (Gao, W. *et al.*, 2021; Oh, H. *et al.*, 2024)

CONCLUSION

The evolution from traditional similarity metrics to learned distance functions represents a paradigmatic shift in how modern recommendation systems approach content retrieval and personalization. Traditional geometric measures like cosine similarity and dot products have established themselves through proven computational efficiency and seamless integration with existing infrastructure, making them attractive for organizations prioritizing scalability and operational simplicity. However, these static approaches fundamentally cannot adapt to the dynamic nature of user preferences, temporal patterns, and contextual factors that significantly influence content relevance in contemporary digital environments. Learned distance functions address these limitations by leveraging neural networks to capture complex, non-linear relationships between users and content. This results in superior precision, enhanced user engagement, and improved content discovery capabilities. The personalization potential of these

Publisher: SARC Publisher

systems, combined with their ability to integrate multiple signal types and adapt continuously through incremental learning, positions them as the future standard for sophisticated recommendation platforms.

Nevertheless, the implementation complexity, substantial resource requirements, and potential for bias amplification demand careful consideration and sophisticated MLOps infrastructure. Organizations must evaluate their specific requirements, user base characteristics, and available resources when choosing between traditional and learned similarity approaches. The emergence of hybrid architectures, multi-modal learning capabilities, and quantum-inspired metrics suggests continued innovation in this domain, promising even more sophisticated solutions for capturing user intent and delivering relevant content experiences.

REFERENCES

1. Lee, H., & Lee, J. "Scalable deep learning-based recommendation systems." *ICT express* 5.2 (2019): 84-88.
2. Marinov, M., Valova, I., & Kalmukov, Y. “Comparative analysis of existing similarity measures used for content-based image retrieval." *2019 X National Conference with International Participation (ELECTRONICA)*. IEEE, (2019).
3. Han, Y., Liu, C., & Wang, P. "A comprehensive survey on vector database: Storage and retrieval technique, challenge." *arXiv preprint arXiv:2310.11703* (2023).
4. Nguyen, B., Morell, C., & De Baets, B. "Large-scale distance metric learning for k-nearest neighbors regression." *Neurocomputing* 214 (2016): 805-814.
5. Tourad, M. C., Abdelmounaim, A., Dhleima, M., Telmoud, C. A. A., & Lachgar, M. "DeepSL: Deep Neural Network-based Similarity Learning." *International Journal of Advanced Computer Science & Applications* 15.3 (2024).
6. Da’u, A., Salim, N., & Idris, R. "An adaptive deep learning method for item recommendation system." *Knowledge-Based Systems* 213 (2021): 106681.
7. Abolghasemi, R., Viedma, E. H., Engelstad, P., Djenouri, Y., & Yazidi, A. "A graph neural approach for group recommendation system based on pairwise preferences." *Information Fusion* 107 (2024): 102343.
8. Gao, W., Fan, X., Wang, C., Sun, J., Jia, K., Xiao, W. & Liu, X. "Learning an end-to-end structure for retrieval in large-scale recommendations." *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. (2021).
9. Chen, R., Liu, B., Zhu, H., Wang, Y., Li, Q., Ma, B., ... & Zheng, B. "Approximate nearest neighbor search under neural similarity metric for large-scale recommendation." *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. (2022).
10. Oh, H., & Kim, C. "Fairness-aware recommendation with meta learning." *Scientific Reports* 14.1 (2024): 10125.

Source of support: Nil; **Conflict of interest:** Nil.

Cite this article as:
Raveendran, N. “Embedding Similarity Metrics in Feed Retrieval Systems - Cosine vs. Learned Distance Functions” *Sarcouncil Journal of Multidisciplinary* 5.7 (2025): pp 196-204.

Publisher: SARC Publisher