



Gated Cross-Attention Matcher for Aligning Course With Relevant KUs

Aashish Pandey, Ram Dantu, Vinh Quach
University of North Texas, Denton, TX



Introduction

We propose a Gated Cross-Attention Module (GCAM) for curriculum–exam alignment by learning robust semantic matching between course knowledge units (KUs) and exam micro-assessments. Raw curriculum artifacts (syllabi, outcome tables) are normalized into KUs containing titles, topics, and learning outcomes. GCAM uses a frozen LLM encoder to generate contextual token states and trains only lightweight LoRA projections plus two interpretable gates: a Token Gate to suppress irrelevant memory tokens and a Head Gate to activate the most useful attention heads per query. The model is trained with contrastive InfoNCE(Information Noise-Contrastive Estimation) loss on supervised text pairs and supports retrieval-based alignment using a KU memory bank.

Problem Definition

A significant research gap exists in the automated alignment of CAE Knowledge Units (KUs) with university-level course content. At present, this alignment is primarily performed manually by instructors, who must examine syllabi, assignments, and related course materials. Such a process is labor-intensive, subjective, and prone to inconsistency. The scientific problem, therefore, is to develop a robust automated framework capable of identifying and modeling semantic correspondence between CAE KUs and course content in a manner that is accurate, consistent, and adaptable to diverse instructional materials.

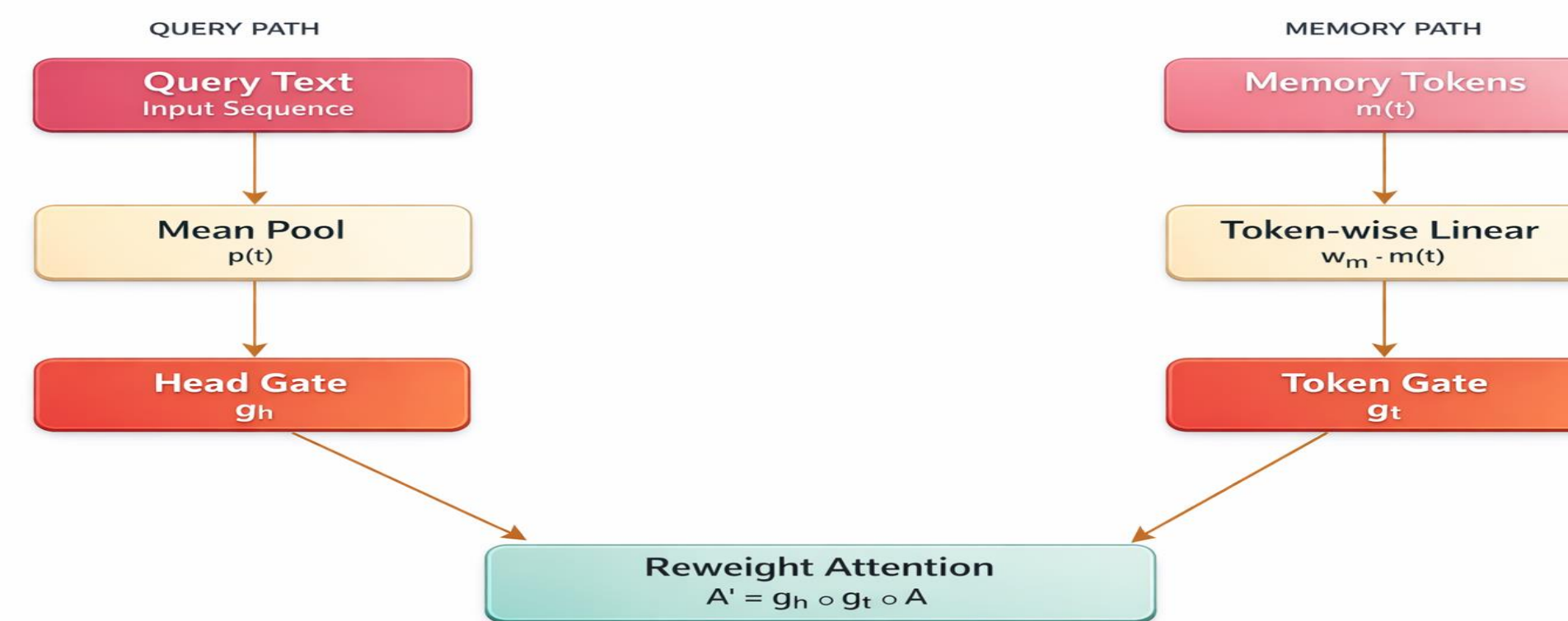
Gates Workflow

Course syllabi and outcome tables describe broad instructional intent. Traditional keyword matching and generic embedding retrieval struggle in this setting due to cross-granularity mismatch, inconsistent phrasing. To address these issues, this project introduces a retrieval-based alignment framework built around a lightweight Gated Cross-Attention Module (GCAM) that learns to match courses to the most relevant curriculum knowledge units.



ARCHITECTURE

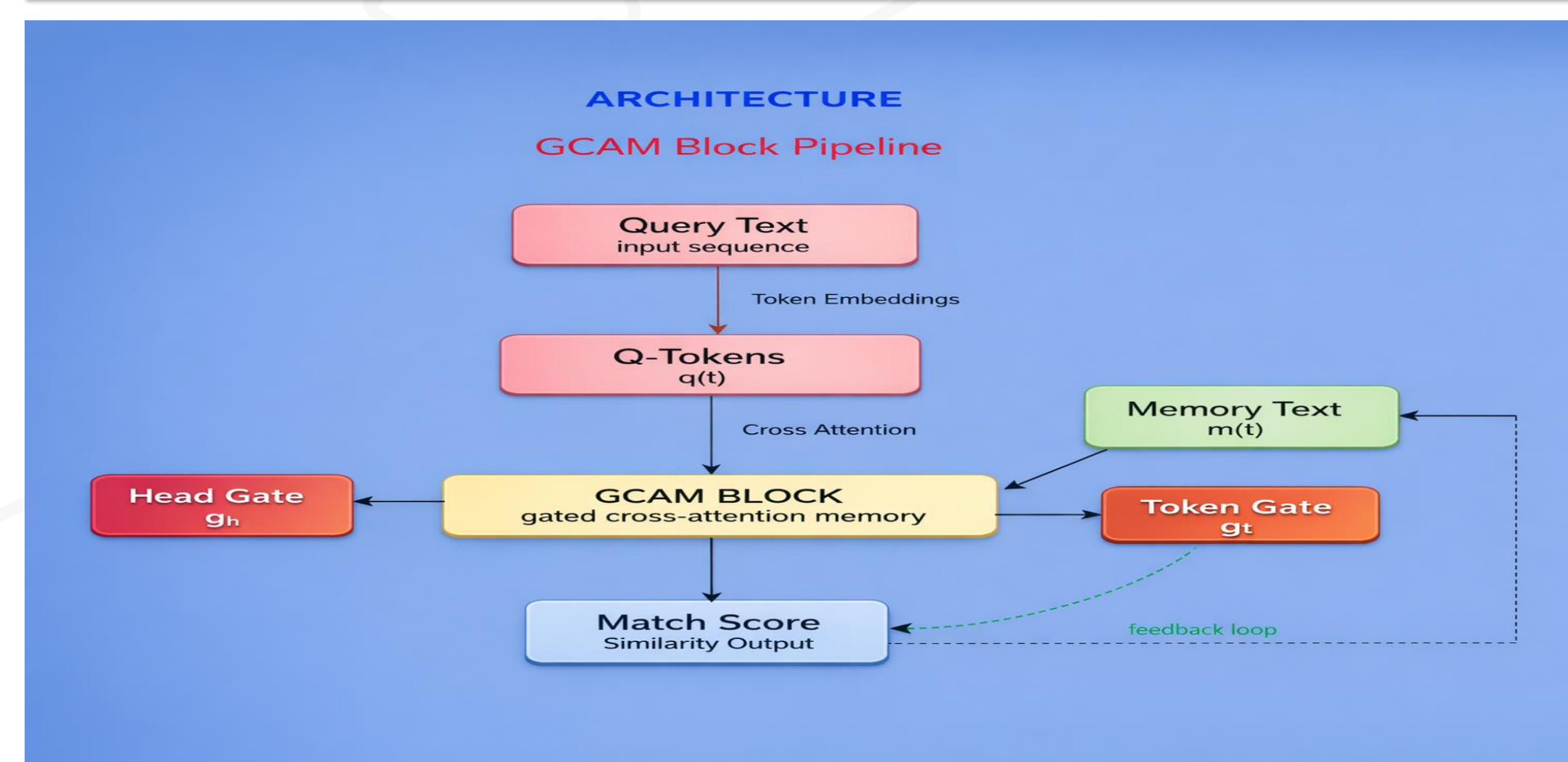
Gated Attention Reweighting



High-level Working of Two Gates; Token Gate and Head Gate

GCAM is an intent-adaptive retrieval framework for aligning educational queries with normalized Knowledge Units (KUs). The pipeline begins by parsing curriculum documents and exam specifications into structured KUs containing titles, topics, and learning outcomes, then constructing supervised text pairs for contrastive training and retrieval. A frozen base LLM encodes both query and memory texts into token states, after which GCAM applies LoRA-based cross-attention to form a query-aware memory summary. It then uses a Token Gate to suppress irrelevant or generic memory tokens and emphasize concept-bearing terms, while a Head Gate selects the most useful attention heads for the query, enabling lexical, semantic, or skill-oriented matching as needed. Together, these components turn standard cross-attention into a noise-resistant mechanism that retrieves the most aligned KUs and outcomes for course titles, topics, and micro-assessments.

Work-Flow



High-Level Working of GCAM with Two Gates

Algorithm

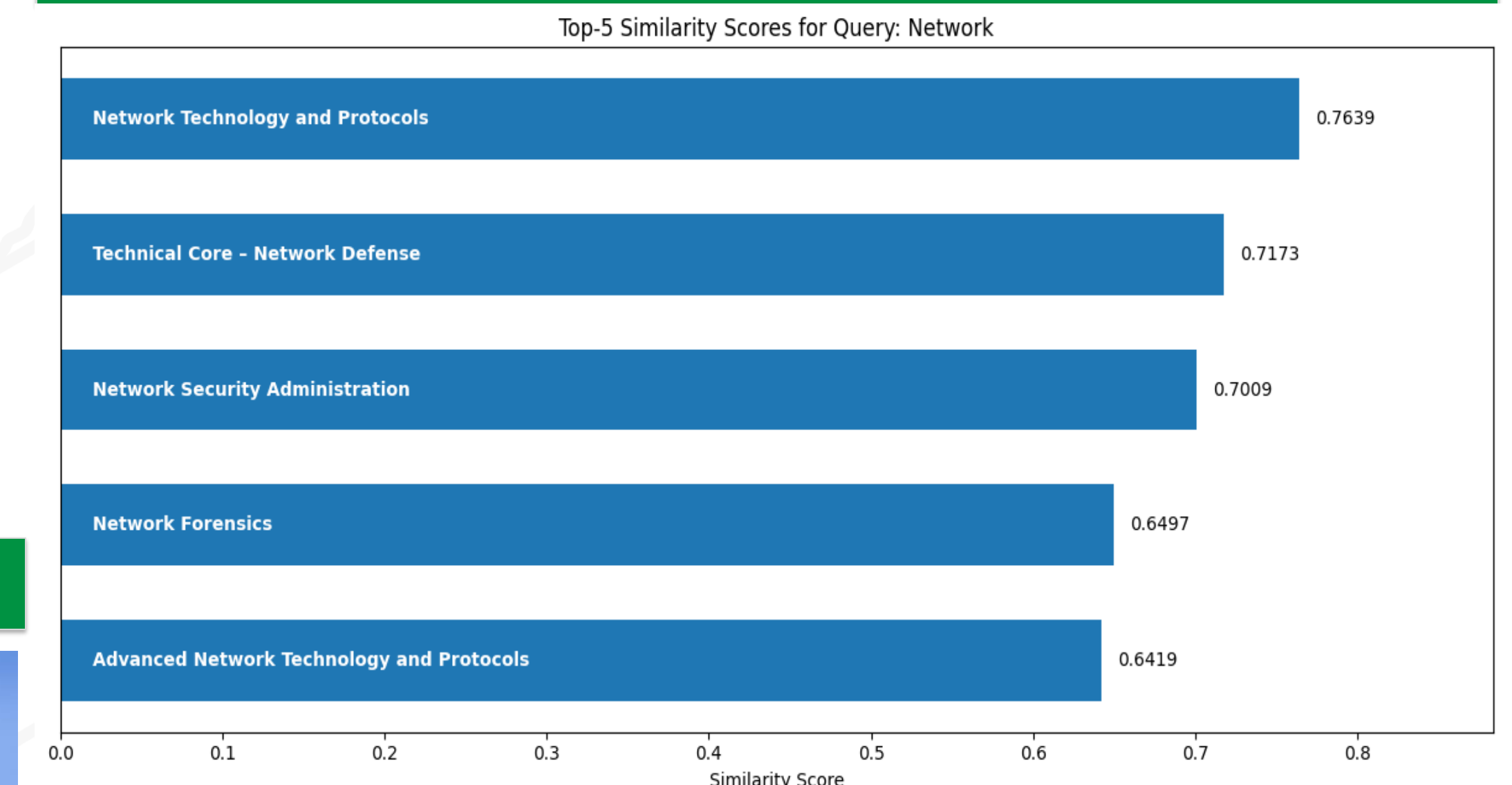
Algorithm 1 GCAM Matching

Input: Query sequence Q , memory sequence M_i

Output: Similarity score s_i

- 1: Encode Q and M_i with a frozen LLM.
- 2: Apply LoRA-enhanced projections and split into attention heads.
- 3: $A_h = \text{softmax}\left(\frac{Q_h K_h^T}{\sqrt{d_h}}\right)$ *cross-attention between query and memory tokens.
- 4: $g_t(j) = \sigma(W_t h_j^M)$, $A'_h = A_h \odot g_t$ *suppress irrelevant memory tokens
- 5: $Y_h = A'_h V_h$, $g_h = \sigma(W_h \cdot \text{mean}(H_Q))$, $\tilde{Y}_h = Y_h \cdot g_h[h]$ *emphasize useful attention heads.
- 6: Merge the gated outputs into a contextual representation.
- 7: $y_Q = \text{mean}(Y)$, $y_M = \text{mean}(H_{M_i})$
- 8: $s_i = \cos(y_Q, y_M)$
- 9: **return** s_i

Results



Conclusion

Our aim is to provide students with an extra evaluation method to assess their skills based on the guidelines of CAE- Knowledge Units. This way they can focus more on the skills they need and become more prepared for industry. Similarly, industry can benefit from this process to find workforce that is more suitable for specific work role by assessing skills of candidates.