Automating Course Articulation: A Deep Metric Learning Framework Using Public Data

Presenter: Mark S. Kim

Duration: 30-45 minutes

1. **Introduction: The Problem with Student Transfer** (~5-7 mins)
   - *Title Slide:* Title, Author, Advisors, University, Date.
   - *The "Transfer Maze":* The current process of determining course equivalency is a manual, inefficient, and intractable barrier for students [1].
   - *The High Cost to Students & Institutions:* This leads to significant credit loss (avg. 43%) [2], delayed graduation, increased financial burden [3], and lower student persistence rates [4].
   - *A Critical Equity Issue:* These barriers disproportionately harm low-income and underrepresented students [5], who are more likely to rely on transfer pathways. Recent transfer growth has been driven by Black and Hispanic students [6].
   - *The Goal & My Contribution:* To design and validate a novel computational framework that automates course articulation using only public data, creating a solution that is:
     - Accurate & Scalable
     - Computationally Efficient
     - Privacy-Preserving
   - *Agenda:* A brief overview of the presentation structure.
2. **Background & Related Work** (~3-5 mins)
   - *The Landscape of Automation:* A brief look at the evolution of automated approaches.
   - *Previous Approaches & Their Limitations:*
     - *Keyword/Statistical (TF-IDF):* Simple but lacks true semantic understanding [7].
     - *Enrollment-Based (course2vec):* Effective but relies on sensitive, private student records, raising significant privacy and generalizability concerns [8, 9].
     - *Direct LLM Classification:* High accuracy potential but operationally challenging due to cost, opacity ("black box"), and prompt sensitivity [10, 11].
   - *The Research Gap:* A clear need for a framework that harnesses the semantic power of modern models without their operational burdens or privacy issues.
3. **A Decoupled Framework for Articulation** (~8-10 mins)
   - *High-Level Architecture:* Visual diagram illustrating the full pipeline from raw text to classification.
   - *Core Principle: Decoupling Representation from Classification.*
   - *Step 1: Deep Contextual Embeddings:* Convert raw course catalog text into rich, high-dimensional vectors using transformer models [12].
   - *Step 2: The Composite Distance Vector (Δ_c):* Our novel feature engineering technique.
     - Combines granular, dimension-specific differences (local information) with a holistic cosine similarity score (global information).
     - Formula: $\Delta_c = (a_1 - b_1, ..., a_k - b_k, \frac{A \cdot B}{\|A\|\|B\|})$.
   - *Step 3: Domain-Specific Fine-Tuning:* Applying deep metric learning to adapt a generic model to the specific language of academia.
     - Objective: Train the model using a Triplet Loss function to create a more discriminative embedding space [13, 14].

- *Step 4: Downstream Classification:* Feed the engineered Δ_c vectors into efficient, traditional ML classifiers (e.g., SVM, Random Forest) for final prediction.

4. **Experimental Setup & Results** (~8-10 mins)

- *Data & Evaluation:* Using the real-world PPM Corpus, partitioned into non-overlapping training and test sets. Primary metric is the F_1-Score.
- *Finding 1: Domain-Specific Fine-Tuning is Critical.*
  - Our fine-tuned model (bge-ft) achieved the highest mean test score and lowest variance.
  - It was *statistically significantly* superior to all off-the-shelf models, including those orders of magnitude larger.
  - **Key Insight:** For specialized domains, targeted adaptation is more effective than sheer model scale.
- *Finding 2: Final Classifier Performance.*
  - All finalist models achieved exceptionally high accuracy (F_1-scores > 0.99).
  - *Support Vector Machine (SVM):* Statistically the most accurate and consistent model.
  - *Random Forest & XGBoost:* Nearly as accurate but an order of magnitude faster and more efficient at inference time.
  - **Key Insight:** A clear trade-off exists between peak accuracy and operational efficiency.

5. **Qualitative Analysis: Beyond the Metrics** (~4-6 mins)

- *Why Do Errors Still Occur?* Aggregate metrics can hide systematic failures [15].
- *Shared Misclassifications:* A high overlap of errors across all models (211 common errors) points to challenges in the data, not the models.
- *Root Cause Analysis (Examples):*
  - *False Negatives:* Caused by semantic divergence (equivalent courses with very different descriptions) or minimalist descriptions. The model correctly sees no textual similarity; the ground-truth label is the issue.
  - *False Positives:* Caused by high topical overlap without true equivalence (e.g., sequential physics courses).
- *The Primary Bottleneck:* The limiting factor for performance has shifted from being **model-centric** to **data-centric**.

6. **Conclusion** (~2-3 mins)

- *Limitations:* Briefly acknowledge boundaries (performance capped by data quality, model generalizability, handling of complex articulations [16]).
- *Future Work:* Highlight key directions (data-centric AI, building a recommendation engine, graph-based methods for complex rules).
- *Summary of Contributions:*
  - Developed a novel, accurate, and scalable framework for automating course articulation.
  - Proved that domain-specific fine-tuning outperforms sheer model scale for this task.
  - Delivered a practical, privacy-preserving tool to reduce administrative burden and foster educational equity.

7. **Thank You & Questions** (~10-20 mins Q&A)

- Final slide with contact information/acknowledgments.

References

[1] Zachary Pardos, Hung Chau, and Haocheng Zhao. “Data-Assistive Course-to-Course Articulation Using Machine Translation”. In: (July 2019). DOI: `10.1145/3330430.3333622`.

[2] United States Government Accountability Office. *Higher Education: Students Need More Information to Help Reduce Challenges in Transferring College Credits*. Report to Congressional Requesters GAO-17-574. United States Government Accountability Office, Aug. 14, 2017. URL: `https://www.gao.gov/products/gao-17-574`.

[3] Leticia Tomas Bustillos et al. *The Transfer Maze: The High Cost to Students and the State of California*. The Campaign for College Opportunity, Sept. 17, 2017.

[4] Stephen Porter. “Assessing Transfer and Native Student Performance at Four-Year Institutions”. In: *39th Annual Forum of the Association for Institutional Research*. June 1999.

[5] The National Task Force on the Transfer and Award of Credit. *Reimagining Transfer for Student Success*. Report to Congressional Requesters. American Council on Education, Mar. 2020. URL: https://www.gao.gov/products/gao-17-574.

[6] National Student Clearinghouse. *College Transfer Enrollment Grew by 5.3% in the Fall of 2023*. URL: https://www.studentclearinghouse.org/news/college-transfer-enrollment-grew-by-5-3-in-the-fall-of-2023/ (visited on 06/30/2025).

[7] Akiko Aizawa. "An information-theoretic perspective of tf-idf measures". In: *Information Processing & Management* 39.1 (2003), pp. 45–65. ISSN: 0306-4573. DOI: https://doi.org/10.1016/S0306-4573(02)00021-3. URL: https://www.sciencedirect.com/science/article/pii/S0306457302000213.

[8] Zachary A. Pardos, Zihao Fan, and Weijie Jiang. "Connectionist recommendation in the wild: on the utility and scrutability of neural networks for personalized course guidance". In: *User Modeling and User-Adapted Interaction* 29.2 (Apr. 2019), pp. 487–525. ISSN: 0924-1868. DOI: 10.1007/s11257-019-09218-7. URL: https://doi.org/10.1007/s11257-019-09218-7.

[9] Sharon Slade and Paul Prinsloo. "Learning Analytics: Ethical Issues and Dilemmas". In: *American Behavioral Scientist* 57.10 (2013), pp. 1510–1529. DOI: 10.1177/0002764213479366. eprint: https://doi.org/10.1177/0002764213479366. URL: https://doi.org/10.1177/0002764213479366.

[10] Federico Errica et al. "What Did I Do Wrong? Quantifying LLMs' Sensitivity and Consistency to Prompt Engineering". In: *ArXiv* abs/2406.12334 (2024). URL: https://api.semanticscholar.org/CorpusID:270562829.

[11] Melanie Sclar et al. "Quantifying Language Models' Sensitivity to Spurious Features in Prompt Design or: How I learned to start worrying about prompt formatting". In: *ArXiv* abs/2310.11324 (2023). URL: https://api.semanticscholar.org/CorpusID:264172710.

[12] Jacob Devlin et al. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2019. arXiv: 1810.04805 [cs.CL]. URL: https://arxiv.org/abs/1810.04805.

[13] Florian Schroff, Dmitry Kalenichenko, and James Philbin. "FaceNet: A Unified Embedding for Face Recognition and Clustering". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2015.

[14] Alexander Hermans, Lucas Beyer, and Bastian Leibe. *In Defense of the Triplet Loss for Person Re-Identification*. 2017. arXiv: 1703.07737 [cs.CV]. URL: https://arxiv.org/abs/1703.07737.

[15] Gabrielle Gauthier-melancon et al. "Azimuth: Systematic Error Analysis for Text Classification". In: Jan. 2022, pp. 298–310. DOI: 10.18653/v1/2022.emnlp-demos.30.

[16] Z. A Pardos, H Chau, and H Zhao. "Data-assistive course-to-course articulation using machine translation". In: *Proceedings of the Sixth Conference on Learning@ Scale*. 2019, pp. 1–10.