PennState

Vincent Trost, Kayla Montgomery, Drew Wham
Teaching and Learning with Technology
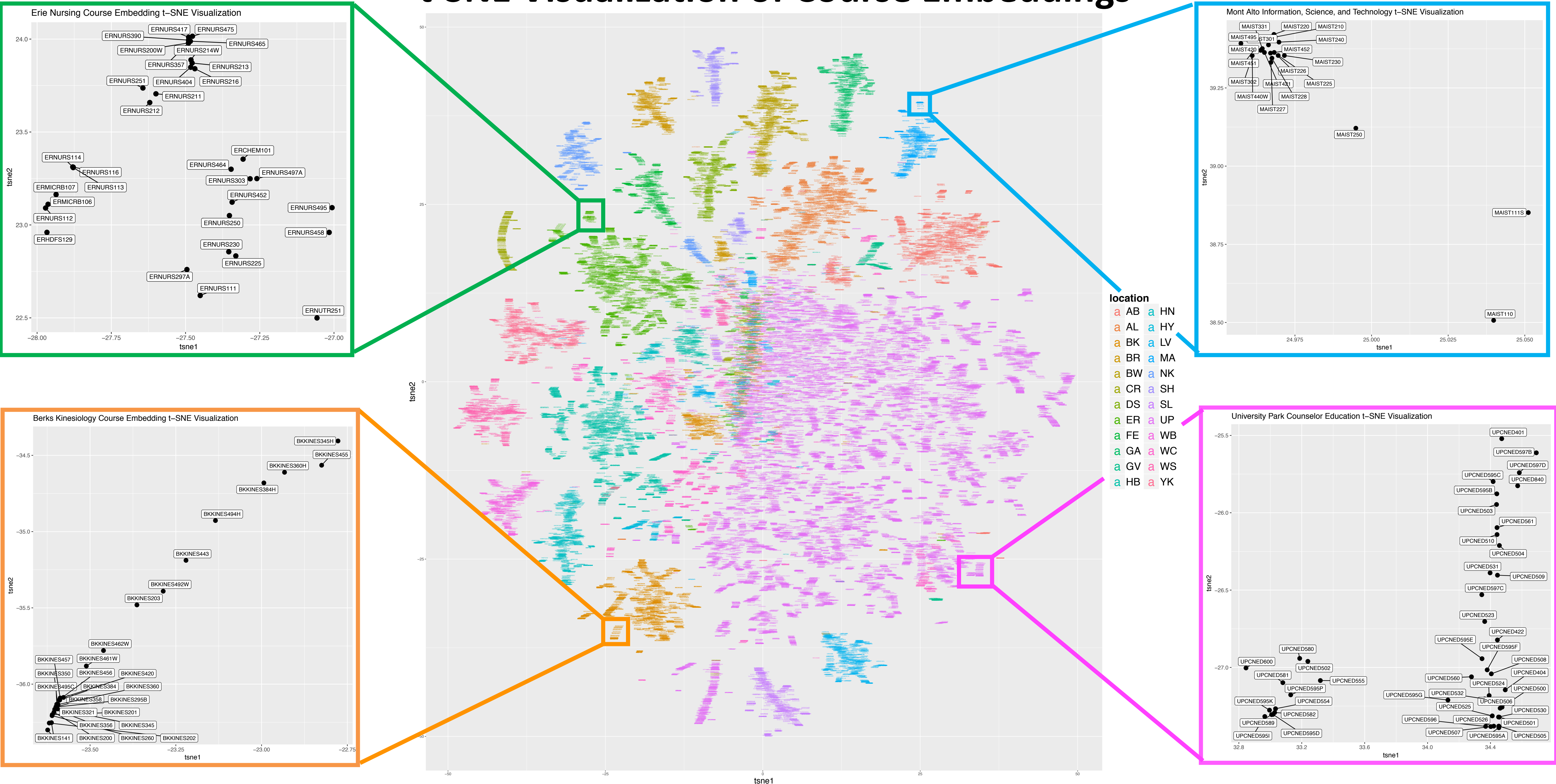The Pennsylvania State University

Course2Vec: Course Embeddings Are Incredibly Useful

Abstract

Embedding methods have been used in Natural Language Processing (NLP) extensively. A number of interesting NLP tasks have benefited from leveraging Neural Networks to produce vector representations of sequentially structured data such as natural text. This approach aims to learn the semantic relationships between words by using the context of the surrounding words. But what if we applied these methods to course sequences where courses took the place of words in the model? Here we used a tool called GenSim to embed courses based on student course history sequences. We then explored these course embeddings to discover if the underlying relationships between courses was well captured by this model. This has the potential to be incredibly useful in enhancing the student experience by impacting the way students choose their courses and advisers advise students.

t-SNE Visualization of Course Embeddings

Most Similar Course

The model yields the similarity measures between Lehigh-Valley's Language and Literacy Education course 401. The model gives us other teaching classes, not just LLED courses. This can help students discover similar classes and it can help advisers identify substitute courses.

Teaching Language Arts in Elementary School	LV-LLED 401	
Music for Classroom Teachers	LV-MUSIC 241	0.9983
Teaching Children's Literature	LV-LLED 402	0.9973
Teaching Reading in the Elementary School	LV-LLED 400	0.9968
The Visual Arts in the Elementary School	LV-AED 303	0.9951
The Young Child's Play as Educative Processes	LV-ECE 479	0.9824

Next Course

The model can be used to predict the probability that a certain course will come after the given sequence of photo, graphic design, and IST courses. Note how it focuses on GD courses. This can help students decide on courses to take in the future.

UP-PHOTO 200 UP-GD 200 UP-GD 201 UP-IST 256 UP-IST 250	Abbreviation	Probability
The History of Graphic Design	UP-GD 202W	0.3435
Typography	UP-GD 201	0.2192
Graphic Design Studio I	UP-GD 200	0.1268
Photo Studio II	UP-PHOTO 300	0.0939
Introductory Design Studio	UP-GD 102	0.0543

Does Not Belong

The model is able to pick out courses that are the least similar to others in the list. The choice of the model is subtle in the first case. Elementary Statistics (STAT 200) is a course taken by many others, not just common STEM majors.

UP-MATH 140 UP-MATH 141 UP-MATH 220 UP-MATH 230 UP-STAT 200 UP-STAT 414	UP-STAT 200
LV-LLED 401 LV-MUSIC 241 LV-LLED 402 LV-LLED 400 LV-AED 303 LV-ECE 479	LV-ECE 479
UP-PHOTO 200 UP-GD 200 UP-GD 201 UP-IST 256 UP-IST 250 UP-PHOTO 300	UP-IST 256