



Siddharth Adelkar, Jenny (Weijie) Jiang

Faculty Advisor: Professor David Bamman, Collaborator: Professor Zachary Pardos (School of Education)

1. Background, Contributions and Datasets

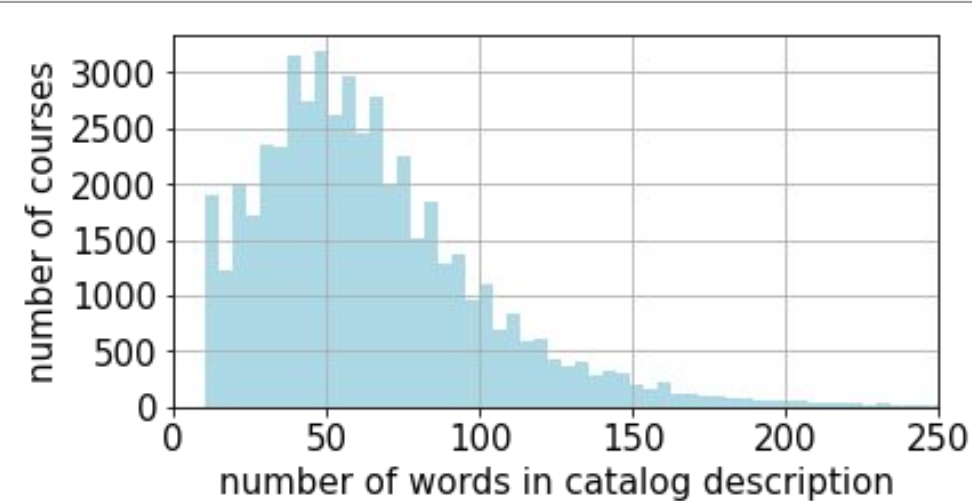
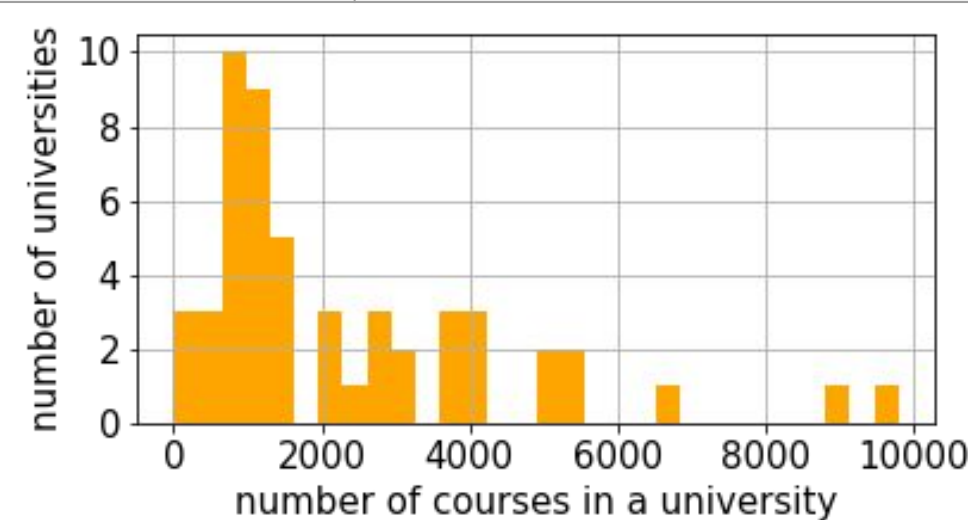
Background: Higher education institutes have been promoting socioeconomic mobility by supporting student transfer from 2-year community colleges to 4-year degree granting universities, such as in the California public post-secondary system and SYS1. However,

- **Course articulation**, *i.e.*, defining which course at one institution will count as credit for an equivalent course at another institution, requires efforts from both the schools, and it is intractable when attempting to manually articulate every set of courses at every institution pair.
- The Community College Research Center reported that the **transferring credit problem** is a major contributing factor to the **dismal rates of graduation** among transferring students.

Goal: To enhance the process of defining and maintaining **course articulations** for SYS1 by leveraging **enrollment patterns** and **course catalog information**.

Datasets: Real-life enrollment and catalog data from a large system of colleges and universities (SYS1) with 58 campuses

Catalog	~120K courses with titles and descriptions
Enrollment histories	~16.7M enrollments of 1.27M students from 2014 Fall to 2019 Spring (anonymous student ID, year, semester, course ID etc.)
Mapping	~128K course articulation pairs (source course, source university, target course, target university)



Contributions:

Social impact: Democratize course articulation by revealing successful transfer pathways to those with no access or social capital.

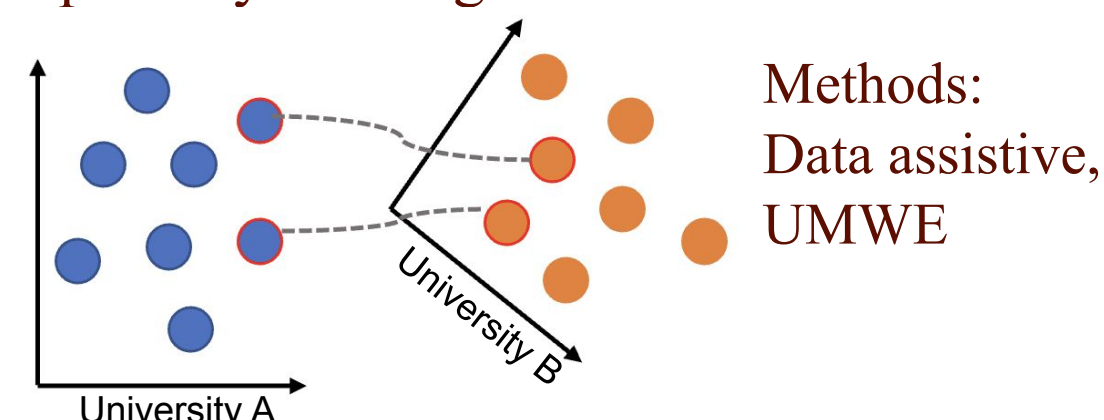
Methodological: We analogize course space to the word space in machine translation and arrive at a universal embedding space.

Practical: Our models will be deployed to the SYS1 system to facilitate course articulation between the 58 campuses.

2. Methodologies

Problem definition: Given a course in university A, find the most similar course in university B. **Two ways of modeling:**

1. Learn course embeddings within universities separately and align them to the same vector space.
2. Learn course embeddings across universities directly

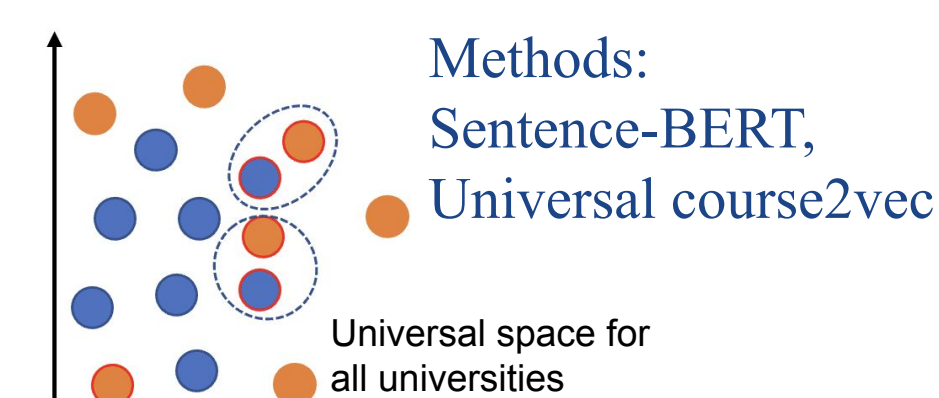
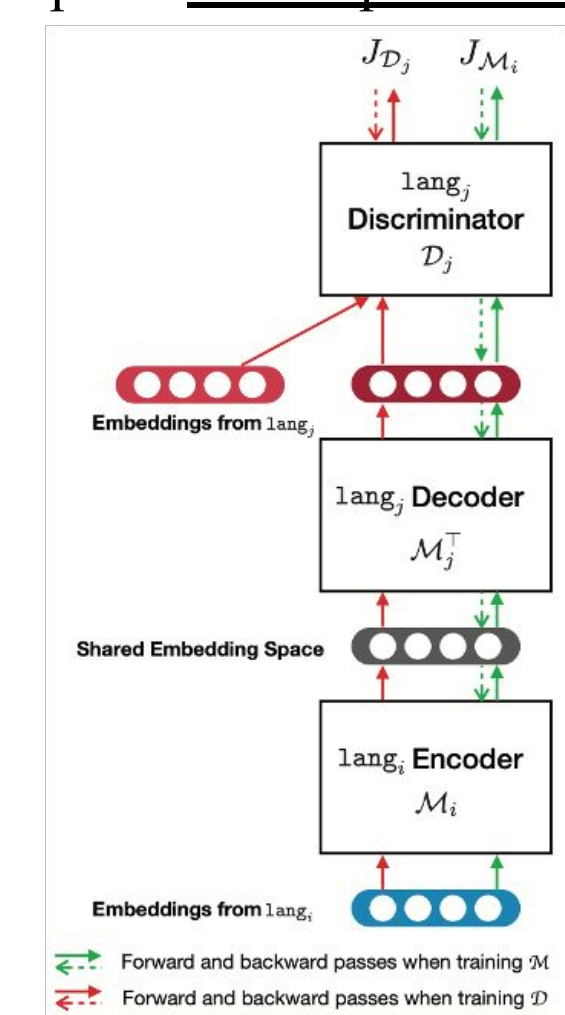


Data assistive (Pardos et al., 2019):

- Learn course embeddings from each university and then learn a linear transformation to map them into the same vector space based on articulation pairs.

UMWE (Unsupervised multilingual word embeddings, Chen and Claire, 2018)

- Use adversarial learning to map all the words in different languages into the same vector space. **We adapt UMWE to course space.**

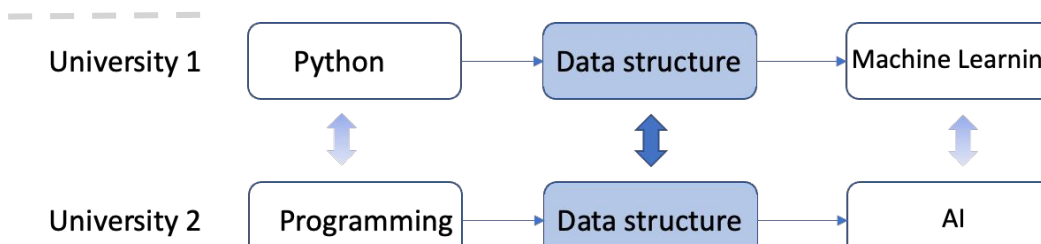


Sentence-BERT (Nils and Iryna, 2019):

- a siamese BERT network to derive semantically meaningful sentence embeddings. **We applied it to course catalog and generated course catalog embeddings.**

Universal course2vec (proposed):

- Assumption: similar courses in different universities might also have similar prerequisites and post-requisites.



The two data structure courses are transferable, which means their prerequisites and post-requisites might also be transferable.

Solution: set a universal course ID for the courses that are transferable, and apply **course2vec** (Pardos et al., 2019) to the enrollment sequences to learn universal course embeddings.

Examples of courses that are transferable in the SYS1 system:

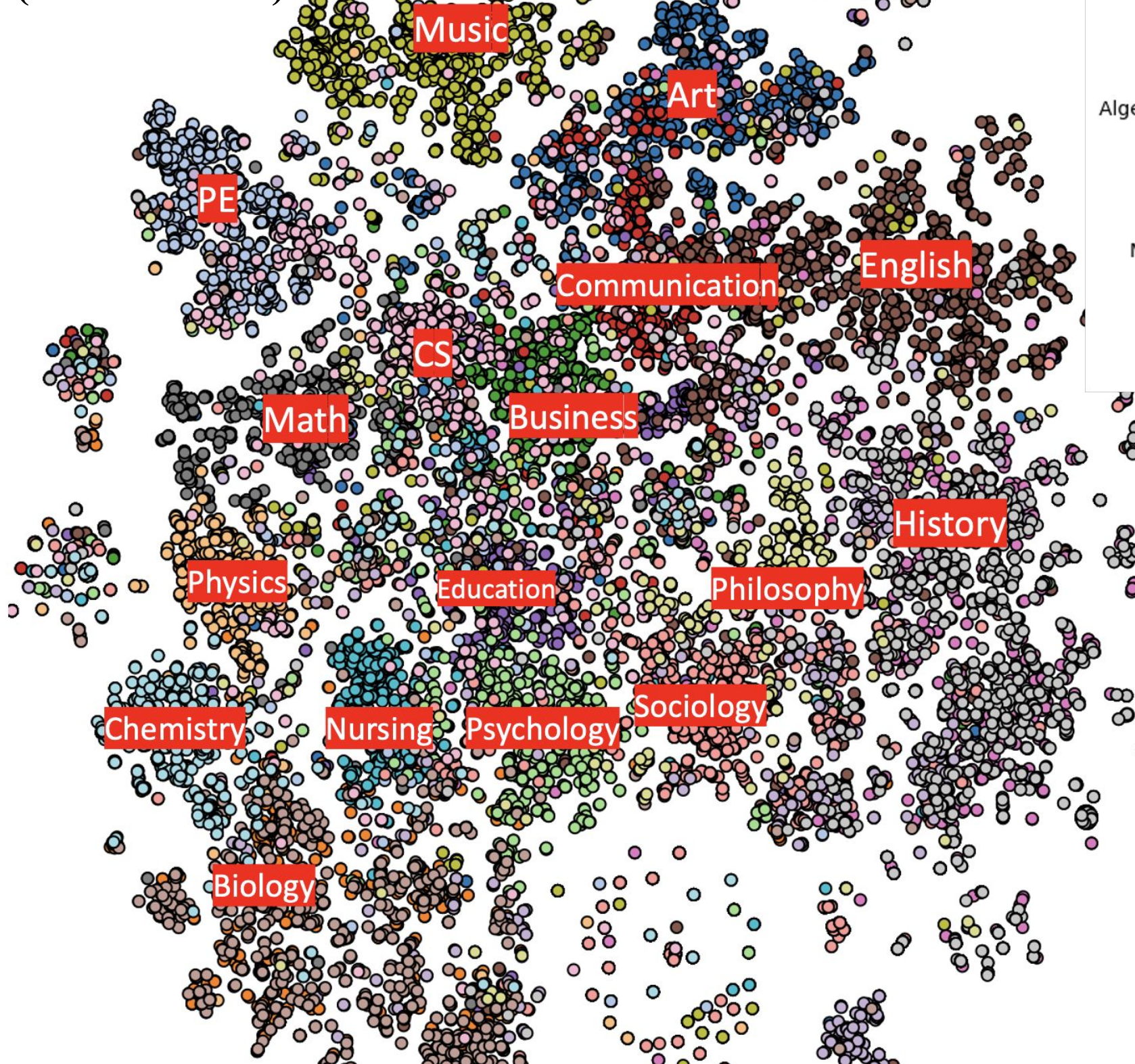
College Algebra
Calculus I
Calculus A
Pre-Calculus
Applied Calculus I

Academic Writing I
College Writing I
Creative Writing
Poetry Writing
Writing Workshop I

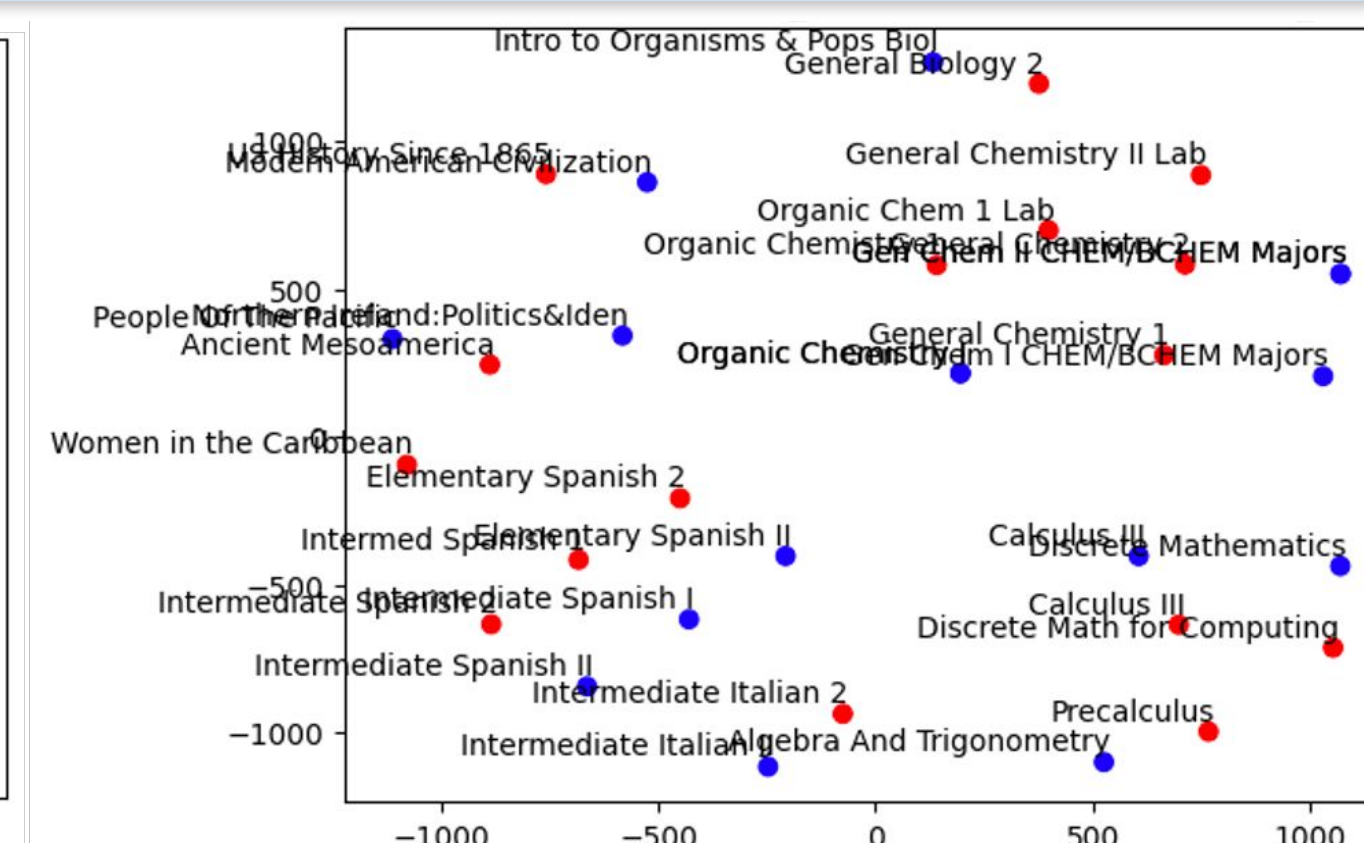
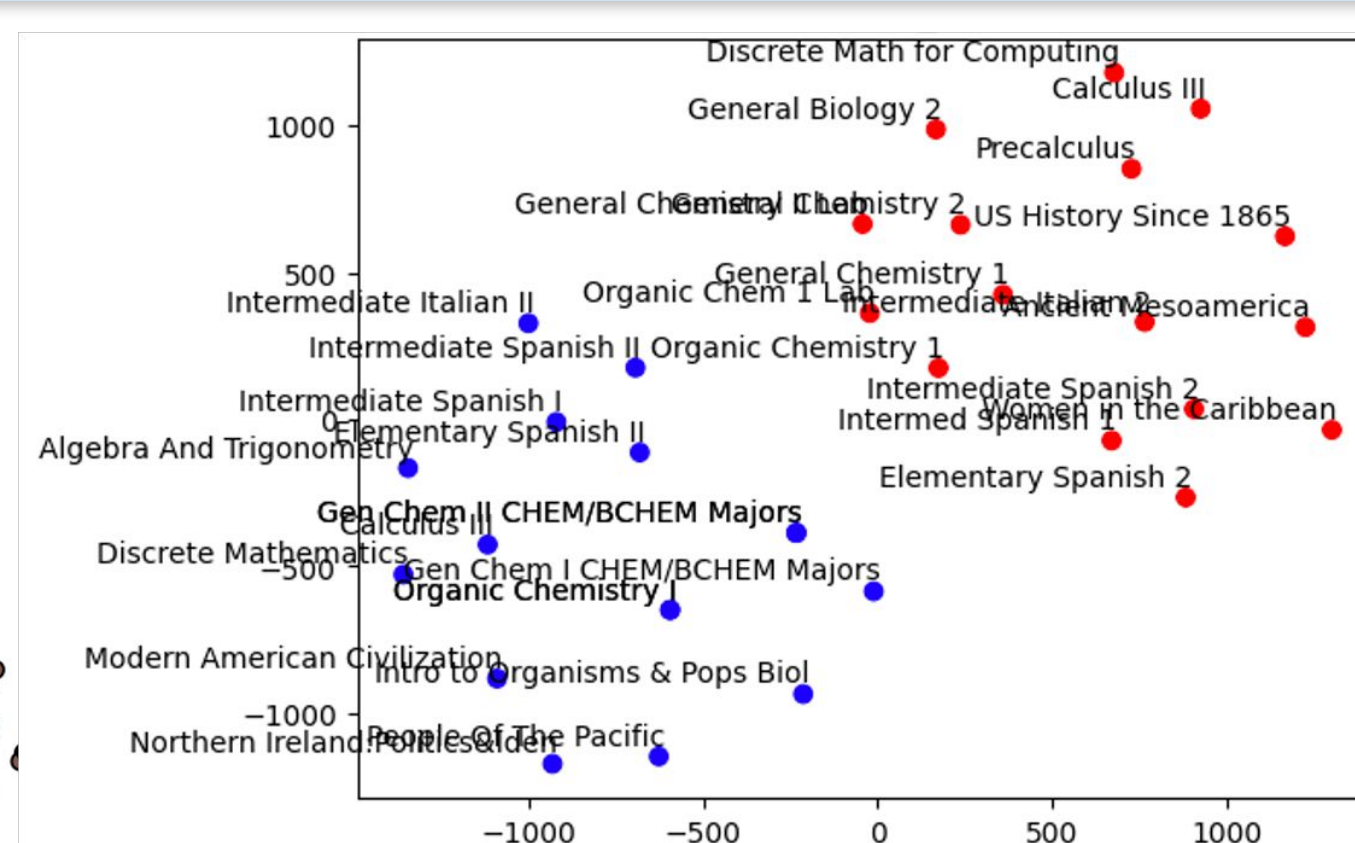
General Chemistry A
General Chemistry I
Elementary Chemistry I
Basic Chemistry
Contemporary Chemistry

3. Experiment Results & Analysis

t-SNE visualization of course embeddings in the SYS1 system (58 universities)

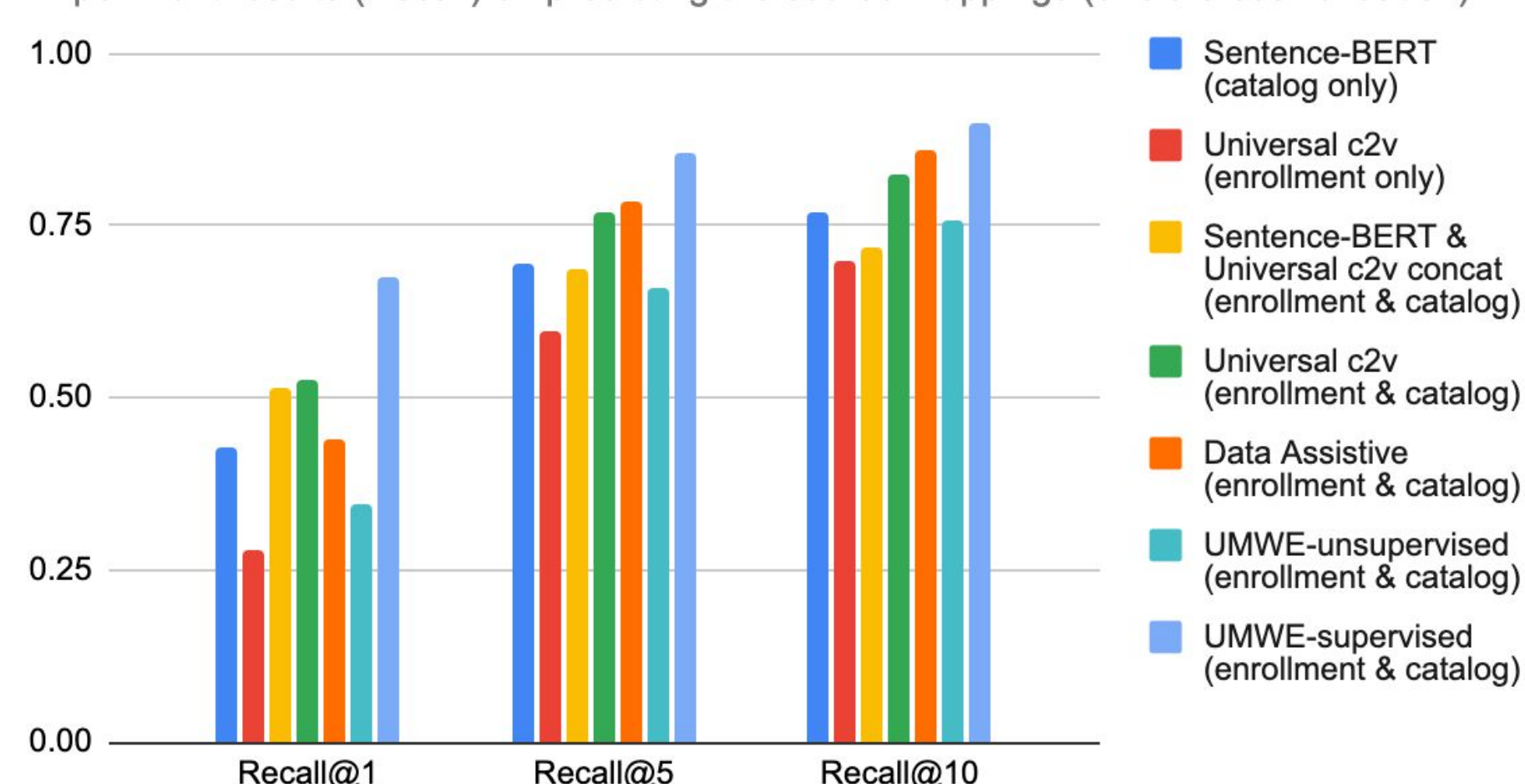


Courses with the same subject are clustered together (the same color). Similar subjects are also closer to each other, for example, Communication is in the midst of Art, English, and Business; Chemistry is closer to Biology and Physics than other social related courses.



The courses in the two universities are initially separated. After the model was trained, the courses are blended to the same vector space where similar courses across universities became closer.

Experiment results (Recall) on predicting the course mappings (5 fold cross validation)



- The UMWE supervised method performed the best among all the methods.

- Combining the catalog and enrollment data together boosted the performance to a large degree.