

# ML-Driven Course Articulation for California Community Colleges



Pranav Gunhal  
pgunhal@ucsb.edu

Annika Damstedt  
acdamstedt@ucsb.edu

Nabeel Nasir  
nabeeln@ucsb.edu

## How can we reduce inefficiencies in CS course articulation while reducing overhead?

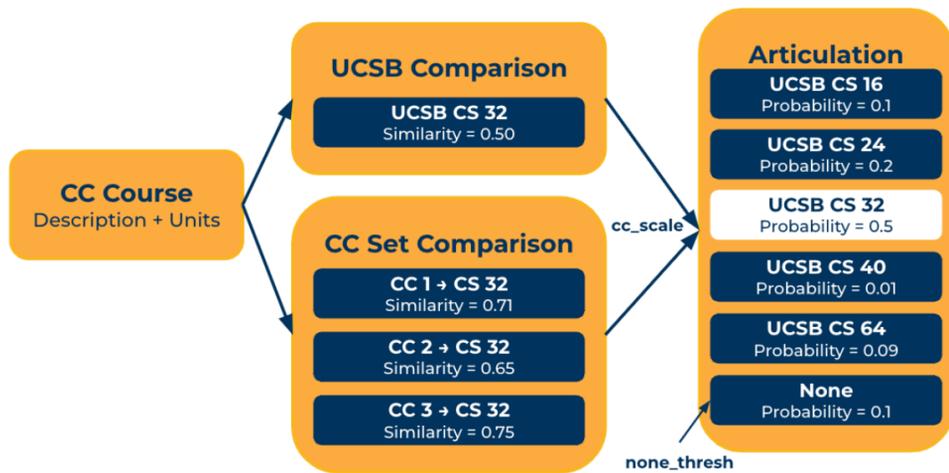


- ▶ One professor determines articulation, which introduces variability and latency
- ▶ Administrative overhead for tracking hundreds of pending articulations
- ▶ Transfer students risk repeating courses, or being placed into courses they are not ready for.

What if we used existing articulations to help predict unknown ones?

## Use set of existing CC articulations for comparison

Compare course description with UCSB reference as well as all of the existing CC articulations for that course.



**cc\_scale** (0-1) - how much the CC-set comparison influences the final score.

**none\_thresh** (0-1) - minimum confidence required to label a course as articulating to no course.

## Existing approaches compare CC→UC in isolation without leveraging existing articulation data

### Course2Vec

*Pardos et al. - PLOS ONE (2020)*

Captures enrollment patterns instead of course content - leverages past student data

### Machine Translation

*Pardos et al. - ACM Learning @ Scale (2019)*

Creates vector mappings of courses across institutions - highly institution specific

### Bag of Words

*Liang et al. - AAAI (2017)*

Captures word overlap in course descriptions - single-institution scope

### Map of Knowledge

*Pardos & Nam - arXiv (2018)*

Uses semantic similarity of course descriptions - misses vocabulary differences

## CC-CC comparison outperforms UC-CC comparison

**91.1%**  
Test Accuracy  
Macro F1 = 0.913  
vs. UC-CC only baseline

**67.9%**  
Test Accuracy  
Macro F1 = 0.629

Validation Accuracy / Macro F1 vs CC Similarity Weight



Best performance near **cc\_scale** ≈ 0.6, indicating higher weight on CC-set similarity than UCSB reference.

(a) Model - with CC-set comparison

	none	cs16	cs24	cs32	cs40	cs64
none	8	2	0	0	0	0
cs16	0	10	0	0	0	0
cs24	0	2	6	0	0	0
cs32	0	1	0	7	0	0
cs40	0	0	0	0	10	0
cs64	0	0	0	0	0	10

(b) Baseline - UCSB similarity only

	none	cs16	cs24	cs32	cs40	cs64
none	0	2	1	6	0	1
cs16	0	9	0	1	0	0
cs24	0	2	6	0	0	0
cs32	0	3	0	5	0	0
cs40	0	0	0	0	9	1
cs64	0	0	0	1	0	9

## Architecture is modeled from real-world articulation

### STACKED BINARY CLASSIFIER

For each UCSB course, a **Logistic Regression** model outputs a similarity score. We choose the highest score, and predict "none" when it falls below *none\_thresh*.

### FEATURES

**Term Frequency-Inverse Document Frequency (TF-IDF) bigrams** - emphasize distinctive terms, down-weight common words.

**CC→UCSB cosine similarity** - compare the CC course description to all UCSB CS courses, compute similarity scores, and pick the closest UC match.

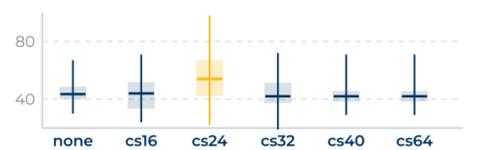
**CC→CC-set similarity** - compare the CC course to CC courses that already articulate to each UC class, and measure fit (ex. max / mean / percentile vs. that set).

### Mean TF-IDF Cosine Similarity Between Classes

	none	cs16	cs24	cs32	cs40	cs64
none	0.063	0.038	0.024	0.030	0.019	0.017
cs16	0.038	0.138	0.049	0.062	0.023	0.029
cs24	0.024	0.049	0.115	0.049	0.026	0.019
cs32	0.030	0.062	0.049	0.129	0.019	0.036
cs40	0.019	0.023	0.026	0.019	0.315	0.020
cs64	0.017	0.029	0.019	0.036	0.020	0.171

Diagonal cells show within-class consistency. There is the most overlap between CMPSC 16/24/32. CMPSC 40 and 64 have the most consistent descriptions in the CC set.

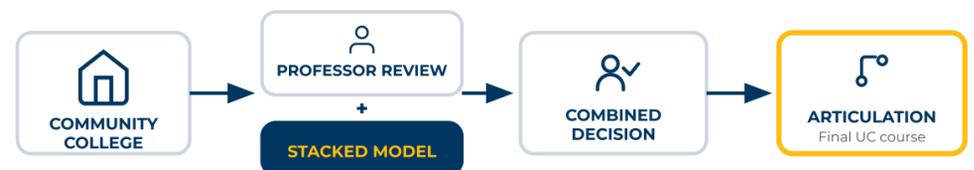
### Avg Description Length by Course



Box/whisker plot of word-count distributions by course.

## Conclusion + Future Work

Our results show that leveraging existing articulations provides valuable data for determining articulations, instead of just direct UC course comparison.



### KEY TAKEAWAYS

- ▶ CC-set comparison increases accuracy relative to UC-CC comparison with only publicly available data
- ▶ Content overlap in sequences of CS courses = route ambiguous cases to faculty

### FUTURE WORK

- ▶ **Syllabus data** will provide info about assessment structure and rigor
- ▶ **More UC data** will help generalize results to other UC/CSU campuses and disciplines
- ▶ **Flag changes** in articulations as courses change at either the UC/CC level
- ▶ **Course sequences** can be added to support two-course CC articulations