Extracting Course Similarity Signal using Subword Embeddings

Yinuo Xu
University of Michigan School of Information
Ann Arbor, Michigan, USA
yinuoxu@umich.edu

Zachary A. Pardos
UC Berkeley School of Education
Berkeley, California, USA
pardos@berkeley.edu

ABSTRACT

Several studies have shown the utility of neural network models in learning course similarities and providing insightful course recommendations from enrollment data. In this study, we explore if additional signals can be found in the morphological structure of course names. We train skip-gram, FastText, and other combination models on these course sequence data from the past nine years and compare results with state-of-the-art models. We find a 97.95% improvement in model performance (as measured by recall @ 10 in similarity-based course recommendations) from skip-gram to FastText, and 80.75% improvement from the current best combination model to the previous state-of-the-art model, indicating that the naming convention of courses (e.g., PHYS_H101) carries valuable signals. We define attributes with which to categorize course pairs from our validation set and present an analysis of which models are strongest and weakest at predicting the similarity of which categories of course pairs. Additionally, we also explore course-taking culture, analyzing if courses with the same demographic features are learned to be more similar. Our approach could help students find alternatives to full courses, improve existing course recommendation systems and course articulations between institutions, and assist institutions in course policy-making.

CCS CONCEPTS

• **Applied computing** → **Education**; • **Computing methodologies** → **Machine learning**.

ACM Reference Format:
Yinuo Xu and Zachary A. Pardos. 2024. Extracting Course Similarity Signal using Subword Embeddings. In *The 14th Learning Analytics and Knowledge Conference (LAK '24), March 18–22, 2024, Kyoto, Japan.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3636555.3636903

1 INTRODUCTION

Course selection is a complex and important decision-making process involving multiple factors, and has long-term implications for students' academic success and career trajectories. Multiple factors come into play, including student characteristics (motivations and abilities), course and instructor features, information availability, and situational considerations [2]. Early course choices, especially for first-year students in fields like engineering, humanities, natural sciences, or social sciences, often predict their eventual majors, influencing career trajectories [1, 4, 19]. Unfortunately, students often lack comprehensive knowledge about available courses, their similarities, and the consequences of their choices [10]. Even with online catalogs, students may struggle to understand courses and their counterparts across different disciplines. Institutions often provide inadequate guidance [21], and students may face enrollment barriers like overcrowded classes. Thus, offering students guidance on course selection, including similar alternatives, is crucial for expanding options and empowering decision-making for academic success. Among various course recommendation approaches, a similarity-based method is essential when students seek alternatives for credit requirements or wish to explore additional courses related to familiar topics. This approach not only aids in creating a diverse curriculum but also serves as a foundation for serendipitous recommendations [16]. Our focus lies in multifaceted notions of similarity, considering factors such as course content, division, or semesters offered.

We aim to address the research gap related to the under-utilization of morphological structures in course names for recommendation tasks. Neural network models, particularly skip grams, have demonstrated suitability for inferring course similarities and predictions (Course2Vec), akin to how skip-gram models (Word2Vec) learn word representations through sentences [15]. Previous work has highlighted the effectiveness of normalized concatenation of the Course2Vec model and course catalog description-based models in ground truth tests of course similarity [16]. In this context, a multi-factor variant of Course2Vec, which learns both course and feature representations (e.g., instructor names and course departments), significantly outperformed Course2Vec. Notably, course catalog descriptions and historical enrollment data prove useful in predicting course time load and difficulty [5]. While similar embedding techniques have been applied in machine translation to assess course similarity across institutions [14], none of these studies have tapped into the morphological meaning of course names in course representations.

We also address the issue of fairness with our model analysis. Fairness in education technology is an emerging concern, as historical bias in existing data could lead to students of marginalized communities being disadvantaged by machine learning algorithms. When predicting the achievement of college students, many grade prediction methods do not accurately predict underachieving students [17], and may underestimate underrepresented demographic groups [24]. The first step to mitigate such bias is to see if it exists in the data. In our study, we investigate if course selections differ by demographic features and if the demographic makeup of courses biases course equivalency tasks.

LAK '24, March 18–22, 2024, Kyoto, Japan

ACM ISBN 979-8-4007-1618-8/24/03…$15.00
https://doi.org/10.1145/3636555.3636903

We assess a novel application of FastText to the Course2Vec task, where it learns embeddings of both full tokens and substrings in a sequence. This approach aims to capture greater course similarity signals by considering substrings of course IDs as they commonly appear in transcripts (e.g., "PHYS_H101"), a feature not included in skipgrams. The resulting sequence-based embeddings are then concatenated with course catalog description embeddings obtained from a pre-trained Sentence-BERT language model and a Bag of Words tf-idf model. Subsequently, we perform a novel attribute analysis for course similarity prediction. This involves defining attributes for ground truth course similarity pairs in our validation set (e.g., similar semester types, both lower division, same course affixes, etc.) and analyzing the performance of skip-gram versus FastText Course2Vec models for these attributes. Our research questions are:

- **RQ1**: Does FastText pick up more course similarity signal than skip-gram from course enrollment data?
- **RQ2**: For what types of similar course pairs do these various models perform best and worst?
- **RQ3**: Do course demographics affect course equivalency performance?

FastText outperforms skip-gram as the Course2Vec model, capitalizing on the morphological meaning of course names for enhanced similarity signals. In our attribute analysis across 23 categories, FastText consistently improves performance, particularly for courses offered in non-summer semesters, and pairs with one STEM and one non-STEM course. Classes with the same demographic features exhibit slightly better performance than those with different features for FastText.

Our work aims to improve course representations for better alternative course recommendations. These recommendations can help reduce student stress related to impacted courses and inform decisions in higher education institutions. Traditionally, faculty have the laborious task of finding suitable alternative courses, often overlooking options in other departments and missing updates. Additionally, for courses in smaller departments with low enrollments, our FastText approach outperforms skip-gram, offering valuable course recommendations. Our methods expand the sets of equivalency pairs, making them more comprehensive. This benefits students by helping them discover alternatives to fully enrolled courses. It also strengthens existing course recommendation systems, empowering advisors to provide personalized recommendations. Furthermore, it supports institutions in formulating effective course policies. Lastly, it enhances cross-institutional tasks like articulation and the identification of similar courses that fulfill requirements.[1].

2 RELATED WORK

Multiple higher education institutions have adopted technologies to support student course selection. One such example is the Open University in Australia, an online educational group that utilizes Personalised Adaptive Study Success (PASS) to personalize students' curriculum planning. To generate curriculum recommendations and feedback, PASS analyzes the student profile, the learning profile, and the curriculum profile[6]. In addition to student-oriented course recommendations, some universities also adopted advisor-oriented recommendations. In collaboration with its counseling staff, the New York Institute of Technology developed its own predictive model to identify at-risk students in need of assistance and assist counselors in their work [12]. Course recommendation approaches like ours could be integrated into higher education institutions to aid both student course selection and advisor advising.

In the realm of course recommendations, various datasets and models have been employed, spanning historical transcript data, student demographics, course information, and enrollment records. Models for grade-aware recommendations include linear singular value decomposition and skip-gram-like log-linear models [11]. Neural networks have emerged for long-term course planning [13] and next-semester recommendations [16]. PLAN-BERT, inspired by masked training and self-attention, supports long-term course plans with consecutive basket recommendations [20]. Previous work has shown that course similarity vectors generated by Course2Vec from enrollment data encode course topics, mathematical rigor, and common student majors [16]. The state-of-the-art (SOTA) course equivalency model that uses enrollment data and catalog data is a combination of multi-factor Course2Vec skip-gram trained on enrollment and BoW tf-idf on catalog [16]. The SOTA course equivalency model uses data from student enrollment, syllabus, and catalog [8]. While integrating diverse data sources has enhanced course recommendations [7, 22], obtaining course syllabus data, used in the SOTA equivalency model, remains challenging due to intellectual property concerns. Given the absence of institutionally sanctioned centralized repositories for course syllabi, we investigate whether our approach can achieve comparable performance using only catalog and enrollment data. FastText has been previously used for Named Entity Recognition in course recommendations to extract skills within a hybrid model [23]. However, none of the prior studies have explored the morphological structure of course names for recommendations. Our study introduces a novel application of the FastText model for equivalency-based course recommendations.

3 DATA

The two datasets used for training models are the enrollment data and course catalog data, both of which are from Fall 2012 to Spring 2021. They are provided by official channels at a large public university in the US. The enrollment data has 6,449,725 rows of records in total in this time frame, with 10,762 unique courses. The course catalog data has 9,116 rows, with 9,116 unique courses.

The data we use to evaluate predicted equivalency is the equivalency validation data maintained by the Office of Registrar, where a course is paired with another one if the student can only get credit for taking one of the courses. The validation test data is also pruned to only include courses in the vocabulary set of enrollment Course2Vec models. Lastly, the validation test data is pruned again to only include pairs that could be predicted by all models, resulting in 732 pairs. Among the validation pairs, 51.09% of the pairs are both STEM courses, 46.72% are both non-STEM courses and only 2.186% of the pairs have one STEM and one non-STEM course.

We define and analyze 23 validation pairs to assess model performance. These pairs are manually engineered based on 5 course categories (student diversity, semester type, division, course affixes, STEM designation), which we consider fundamental units

[1] code is available at https://github.com/CAHLR/representation_presenter

distinguishing various courses. By encompassing both student demographics and intrinsic course properties, these categories ensure that models learn beyond surface-level naming conventions, capturing student course-taking behaviors. The broad categories are further divided into smaller course-level sub-categories, resulting in 23 pair-level categories. For instance, the division type category is subdivided into lower division, upper division, and graduate division, generating pair-level categories like "both lower div," "both upper div," "both grad div," and "different divisions." We exclude 4 categories (different STEM/non-STEM, both cross-listed, different divisions, and both grad divisions) as they constitute less than 5% of the validation test set, rendering their performance non-generalizable. The analyzed attribute categories and their definitions are:

- Diversity and distribution: A course has major diversity if the students enrolled represent more than 27 (median) different majors. A pair of courses have similar distribution if they have at least 3 (median) majors in common among each course's top 10 majors.
- Semester types: Overlapping semesters occur when two courses are both offered in regular semesters (Spring, Fall semesters) or both in Summer semesters.
- Divisions: Courses share the same division if both are lower division (course number below 100), upper division (course number between 100 and 199), or graduate division (course number at least 200).
- Cross-listing and affixes: A course has an affix if it includes the prefix "C" (indicating cross-listing), suffix "AC" (representing the "American Culture" breadth requirement), or other suffixes (e.g., "A", "B", "C" denoting an ordered sequence of courses). Cross-listed courses are those offered jointly by two or more departments, identified by the prefix token "C" (e.g., "Data Science C100" cross-listed with "Computer Science C100"). In our categorization, "cross-listed" means either pair contains the token "C," signifying a course that was not cross-listed before but becomes cross-listed with another department later.
- STEM: Courses are categorized as STEM or non-STEM based on The U.S. Department of Homeland Security (DHS) STEM Designated Degree Program [2].

Preprocessing of enrollment data before model training involve grouping by students and then sorting each student's sequence by semester. Within the semester, the order of courses is randomly shuffled. Next, the Course2Vec model is trained on these class sequences. The course catalog description data used to train the Sentence-BERT model is not pre-processed, while it is for the Bag of Words model by removing stop words, punctuation, and generic sentences across descriptions, word lemmatization and stemming, and tokenizing the bag-of-words in each course description.

We apply three embedding models to course enrollment sequences. Skip-gram and FastText were originally conceived to be applied to natural language, while Multi-factor Course2Vec [16] is a variant on the skip-gram but was first applied to course sequences. Among these three models, FastText had not yet been applied to learn course embeddings from enrollment sequences. We also apply two models to course catalog descriptions. A pre-trained Sentence-BERT model and a Bag of Words (tf-idf) model are used to embed course catalog descriptions. Lastly, different combinations of course sequence embeddings and catalog description embeddings are concatenated.

4 MODELS

4.1 Embedding Models: Skip-gram, Multi-factor, and FastText Course2Vec

Similar to Word2Vec, the skip-gram Course2Vec model represents courses by treating an enrollment sequence as a sentence, considering each class in the sequence as a word [16]. For instance, an enrollment sequence like [Molecular & Cell Biology 160, English 100, Statistics 134] is analogous to a sentence. The Multi-factor Course2Vec enhances Course2Vec by incorporating user-defined features such as instructors and academic departments. In the previous study, the multi-factor model outperformed Course2Vec in recall@10 within the equivalency set. We apply a Multi-C2V model with the same factors to our enrollment data. Course names are morphologically rich, encompassing the department name, course affixes, and course number. Affixes include prefixes like "C" in "Statistics C140" denoting a course offered jointly and suffixes like "A" or "B" indicating courses taken in sequence. In our dataset's institution, course numbers below 100 denote lower-division courses, 100 to 199 signify upper-division, and above 200 indicate graduate-level courses. Skip-gram, however, doesn't capture these naming conventions, assigning a distinct vector to each course. FastText, representing words as character n-grams and accommodating out-of-vocabulary words, may enhance skip-gram's course representations by considering both semantic and morphological meanings [3].

4.2 Course Catalog Models: Sentence-BERT and Bag of Words

Sentence-BERT is a state-of-the-art BERT model that derives semantically meaningful sentence embeddings that can be compared using cosine-similarity [18], reducing the time to find the most similar pair. Here we use Sentence-BERT to obtain the embedding representations learned from course catalogs. Bag of Words represents a piece of text into a vector of real numbers. Tf-idf weighting gives a larger value for words that are rare in the whole document but frequent in a document. We choose to use the tf-idf weighting as it produced the best-performing course catalog embedding in a previous study [16].

4.3 Combination Models

To enhance Course2Vec representation, we concatenate the Sentence-BERT representation of the corresponding course catalog to the Course2Vec embedding, normalizing the embeddings before concatenation. Similarly, we concatenate the normalized course embedding from the FastText model with the catalog embedding from Sentence-BERT. A modification of the multi-factor Course2Vec model, Multi-C2V FastText, is created by vector summing the feature and FastText course embeddings. Lastly, Multi-C2V FastText + Sentence-BERT is a model generated by vector summing the feature embeddings and FastText course embeddings, concatenated

[2] https://www.ice.gov/sites/default/files/documents/stem-list.pdf

with the Sentence-BERT embedding of the catalog descriptions. Multi-C2V skip-gram + BoW tf-idf is the previous state-of-the-art course equivalency model, included as a baseline for comparison.

4.4 Model Training and Evaluation

We employ a validation set consisting of 732 course credit equivalency pairs, defined by faculty. To expand the set's size, we evaluate pairs bidirectionally (A predicting B and vice versa), resulting in 1,464 pairs. Following established metrics from a previous study, we calculate recall@10 for equivalency validation pairs. For each pair, we rank other courses based on the cosine similarity of their vector representations and calculate recall@10 based on the rank of the second course. Ten-fold cross-validation is then utilized to optimize model hyperparameters. The 1,464 pairs are split into 10 folds. Then, within each phase of the cross-validation, 80% of the pairs are used to find the best training hyper-parameters, which are then used to create a model to evaluate on rest of the 20%. The ranks from each fold are aggregated to calculate overall recall@10. Grid search is conducted on hyperparameters for both skip-gram and FastText, including min count: [10, 20, 30, 40, 50, 60, 70, 80, 90], window: [2, 3, 4, 5, 6, 7, 8, 9], vector size: [200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310], sample: [1.e-05, 2.e-05, 2.e-06], alpha: [0.01, 0.02, 0.03, 0.04], min alpha: [0.0001, 0.0003, 0.0005, 0.0007], and negative: [10, 15, 20, 25].

5 RESULTS

5.1 RQ 1: Does FastText pick up more course similarity signal than skip-gram from course enrollment data?

Table 1 shows the recall@10 for the models trained on the enrollment data. Among course embedding models, FastText performs significantly better than skip-gram. For the enrollment models, there is a 97.95% increase from skip-gram to FastText. The best model (FastText + SBERT) improves on the SBERT model by 10.81%, and the state-of-the-art model (Multi-c2v skip-gram + BoW tf-idf) by 80.75% . Thus FastText does pick up more course similarity signals than skip-gram for the enrollment data.

5.2 RQ 2: For what types of similar course pairs do the course2vec models perform best and worst?

We investigate if there were categorical areas in which FastText was making improvements in recall. We conduct an attribute analysis, where we calculated the recall of skip-gram model vs. FastText model for 23 attributes as described in Section 3. The results comparing recall of FastText model vs. skip-gram as the Course2Vec sequence embedding model for course pairs with different attributes are summarized in Fig. 1. FastText outperforms skip-gram in all categories. The top 3 categories that FastText has the greatest advantage in are “neither offered in summer", "neither diverse, similar distributions", and "same course affixes". Additionally, 8 out of 23 categories have low frequency of occurrence in the enrollment data (highlighted yellow in Fig. 1). FastText has an advantage over skip-gram in all categories that have a low frequency of occurrence. We also apply the same analysis for FastText which is trained on enrollment data vs. SBERT which is trained on catalog data in Fig. 2. FastText performs better than SBERT in 6/23 categories, the top categories for FastText being "different departments", "different divisions", and "different course numbers". The top categories for SBERT are "same course number", "different course affixes", and "neither diverse, different distributions".

5.3 RQ 3: Do course demographics affect course equivalency performance?

There has been evidence of algorithmic bias concerning race in grade prediction models trained on historical enrollment [9], meaning that there are different patterns of course selection by race. To rectify racial biases in models, we should first identify where they exist. Based on the assumption that students of the same culture (here we use race, gender, and parent income class as proxies) may share goals and choose courses more similarly than those of different cultures, we hypothesize that the course selection behaviors of students of the majority demographic affect the embedding representations of courses. Assuming there exist different cultures of course-taking that manifest in course embeddings, then course pairs with the same majority demographics should be more accurately predicted than those with different majority demographics. Courses can have multiple majority races and genders, meaning that there is an equal number of students in these categories (typically exist for courses with a small number of students). We divide the continuous variable parent income into 2 categories: below overall median income, and above or equal to median income. Courses with the same demographic features have slightly better performance. The differences between recall@10 of courses that have the same demographics and different demographics (for gender, race, and income) are shown in table 2. We see that course pairs with the same demographic features perform slightly better than those with different features for FastText. Skipgram and SBERT are less sensitive to demographic features.

6 DISCUSSION

FastText provides an 80.75% improvement over the previous SOTA combination model utilizing enrollment and catalog data [16]. Furthermore, compared to a previous SOTA model using multiple data sources (student enrollment, course syllabus, and catalog data) [8], FastText achieves a similar boost in performance using only enrollment and catalog data. While the previous study showed a 49.5% increase from multi-C2V to the best combination model (multi-C2V on enrollment + BoW on syllabus and catalog), the percent improvement from FastText to the best combination model (FastText on enrollment + SBERT on catalog) is 47.6%. Given the challenge of obtaining syllabus data, our approach offers an effective alternative. The strength of our subword approach is particular to how course IDs are structured in our dataset, but it can be extended to different institutions. Most higher education institutions use similar course token conventions, even if expressed differently, to distinguish various course categories. For example, distinguishing between upper and lower-division courses often relies on course numbers. It's this regularity in course tokens that likely makes FastText effective. Given a small set of preexisting course articulation pairs between two institutions [14], FastText could be used to map courses across

Table 1: Equivalency validation results

Model	skip-gram	FastText	BoW tf-idf	SBERT	**FastText + SBERT**	Multi-c2v skip-gram	Multi-c2v skip-gram + BoW tf-idf	Multi-c2v FastText	Multi-c2v FastText + SBERT
Recall@10	0.2008	0.3975	0.5092	0.5295	**0.5867**	0.2891	0.3246	0.3876	0.5673

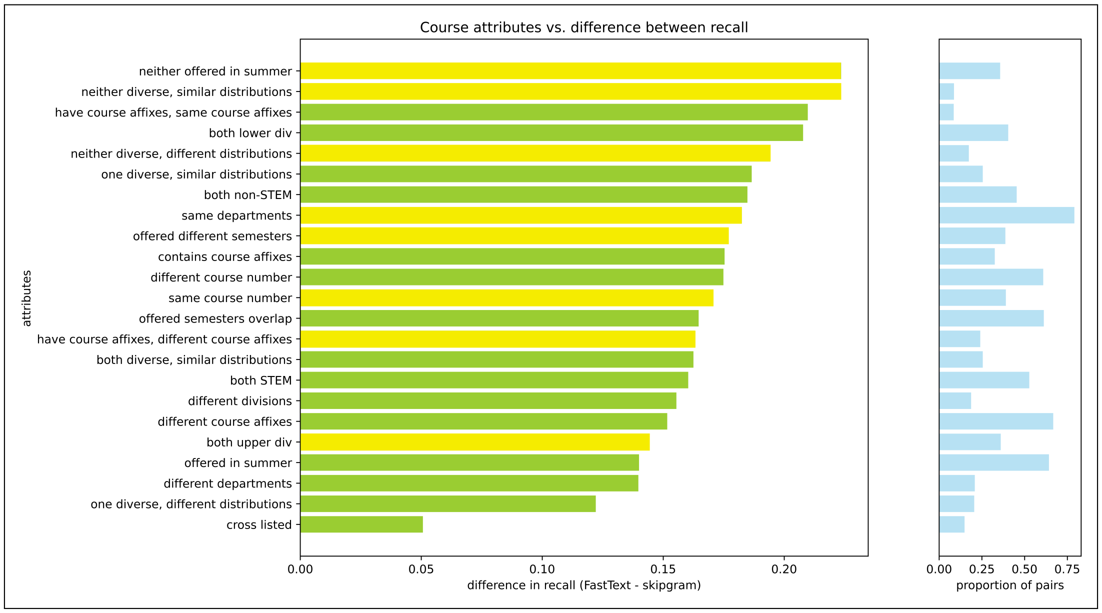

Figure 1: Difference in Recall@10 between skip-gram and FastText, ordered by the descending FastText advantage. The yellow bars indicate categories that have below median frequency of occurrence in the enrollment data. The plot on the right shows the proportion of validation pairs for each category.

Table 2: Recall@10 of course pairs with the same demographic (gender, race, income) features minus recall@10 of pairs with different features.

	gender	race	income
diff in skip-gram recall	0.0159	-0.0309	0.00489
diff in FastText recall	0.0451	0.0381	0.0329
diff in SBERT recall	0.0159	-0.0309	0.00490

institutions. Machine translation or other techniques could then be used to align the spaces, which could identify which conventions represent the same categories (e.g., upper/lower div).

FastText's versatility in capturing similarity signals from diverse sub-words within course names, including token-level details like department and course number, makes it outperform skip-gram across all categories, even in instances with low occurrence frequencies. This is important for articulating courses with low enrollments, broadening transferable course lists, and expanding pathways. In the top 10 categories (Fig. 1), 7/10 show above-median proportions for pairs with the same department and course numbers, with 7/10 featuring below-median occurrence frequencies. Course affixes minimally impact recall (median proportion with the same affix is 9.150%), except when all pairs share identical affixes. Analyzing the bottom 10 categories, 7/10 have proportions below the median for pairs with the same department, and 8/10 have proportions below the median for pairs with the same course numbers. FastText prioritizes course numbers and departments over affixes, likely influenced by its character n-grams' default length (3-6). While length 3 captures most course numbers and length 6 captures most department names, course affixes often merge with other sub-words, making them insufficient for differentiation between courses. Disparities in the top and bottom categories of SBERT's advantage over FastText may be attributed to the semantics of course descriptions and the morphology of course names: courses with "different departments", "different divisions", and "different course numbers" are more morphologically similar than semantically.

FastText performs better in predicting course pairs that share the same majority demographic attributes (race, gender, and parent income), showing an average difference of 9.34%. STEM courses also exhibit a 19.25% improvement in FastText predictions when compared to non-STEM courses. Among pairs that are both STEM, 43.58% have a majority of Asian students, 31.28% have a majority of male students, and 31.55% have a majority of students with parent income at or above the median. This suggests that course

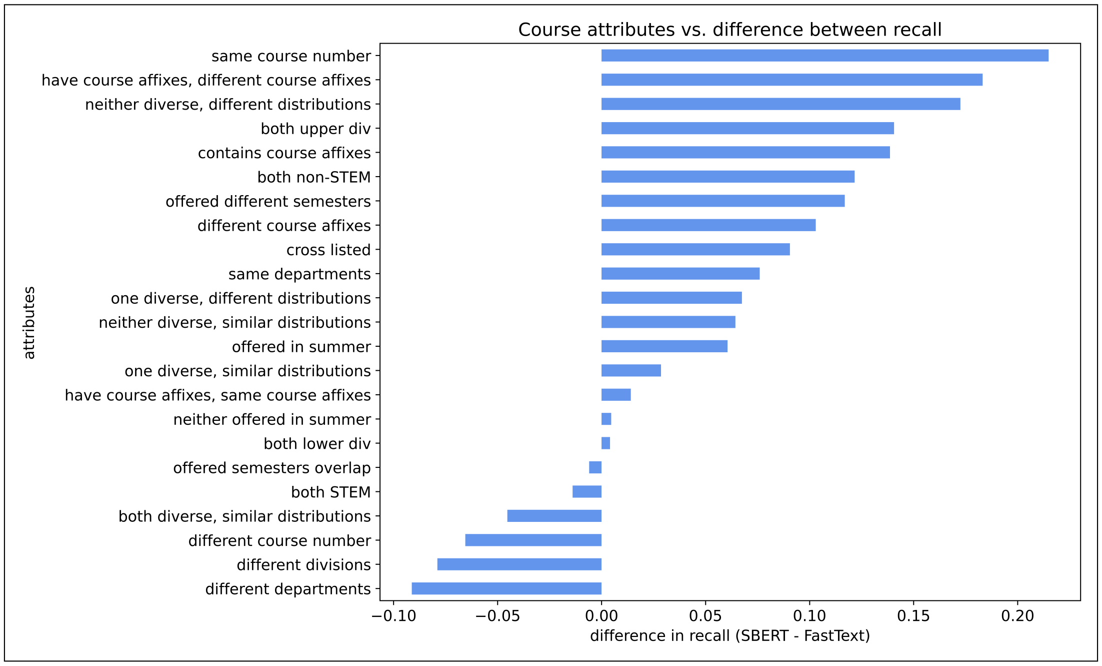

Figure 2: Difference in Recall@10 between SBERT (based on catalog data) and FastText (based on enrollment data), ordered by the descending FastText advantage.

embeddings could be influenced by the bias of the majority demographic, resulting in enhanced recommendations for the majority demographic — a consideration for future fairness work.

This work has several implications for student success and creating personalized learning experiences. By providing more accurate alternative recommendations, institutions could help alleviate student stress about getting into impacted courses. Additionally, course alternative recommendations could also provide students with other possible classes to take if they wish to explore a field further. Moreover, this information can be integrated into institutional policy-making, influencing decisions related to curriculum development, resource allocation, and course articulation between academic departments. Rather than relying solely on faculty members to manually identify equivalent or alternative courses, institutions can leverage model-generated insights to streamline this process. This not only enhances the efficiency of administrative tasks but also ensures that a more diverse set of alternatives, including those from different departments and low enrollments, is considered in the decision-making process. Additionally, institutions can leverage the expanded list of similar courses to inform curriculum design and guide the development of more engaging course sequences for students.

7 FUTURE WORK AND LIMITATIONS

Several future improvements could be pursued in the modeling techniques. Ensemble and fusion methods could be used to take advantage of the strength of each model, minimize the bias in models, and achieve better performance. To further explore more accurate representations of sub-words, which capture course equivalencies significantly better than skip-gram course embeddings, self-attention models could be brought to bear. Furthermore, since our task utilizes data spanning 9 years, there could be multiple versions of the same course, and the subject matter of a course could drift over time. Future work could focus on incorporating temporal information by concatenating course tokens with the academic year, learning the contextual embeddings for various versions of a course, and performing word sense disambiguation. Extending our demographic analysis to include age and first-gen status could give additional insights biases in equivalency estimation and courses differ from those perspectives. To test whether there is a correlation between the culture of course-taking with demographics, we could use adversarial debiasing to remove the cultural features from the course embeddings [25]. This is based on the perspective that the cultural dimension should be treated as an undesired bias, but these demographic attributes might carry useful signals as well. A limitation of our study is that a FastText model trained on the courses of a specific institution might not be generalizable to other institutions, as the effects of FastText may differ with different naming conventions. In future work, to investigate the generalizability of FastText across different course name conventions, we will also data mine from FastText representations to infer universal tokens that have meanings across departments or institutions. Additionally, future work could also consider factors such as course difficulty and learner prior knowledge factors to enhance the recommendations made by a similarity-based approach.

8 CONCLUSION

Our study shows that FastText (Course2Vec) can improve course representations significantly from skip-gram (Course2Vec) and multi-factor Course2Vec models by taking advantage of the morphological meaning of course names as similarity signals: there is an improvement of 97.95% from skip-gram to FastText, and an improvement of 80.75% from the previous state of the art combination model. Our FastText method can provide more accurate course alternative recommendations while using fewer sources of data compared with the current SOTA model. We also present an analysis framework based on 23 different course attributes to identify which models perform best on which types of course pairs. We conclude that FastText performs better than skip-gram in all categories. FastText has the greatest advantage over skip-gram for courses that are offered in non-summer semesters, have similar major distributions, and share the same course affixes. FastText has the greatest advantage over Sentence-BERT for courses with different departments, divisions, and course numbers, while Sentence-BERT has the advantage when comparing courses with the same course number (but different departments), different affixes, different majors, and upper division courses. Lastly, we find that courses that have the same demographic features exhibit a slight advantage over those that do not for FastText, while skipgram and SBERT are less sensitive to such features, a consideration for future fairness work.

ACKNOWLEDGMENTS

This work was funded in part by a grant from MDRC/Ascendium Education Group (Award #1460-UCB-2000-01).

REFERENCES

[1] Ban Cheah Anthony P. Carnevale. and Andrew R. Hanson. 2015. The economic value of college majors. *Georgetown University* (2015).

[2] Elisha Babad. 2001. Students' course selection: differential considerations for first and last course. *Research in Higher Education* 42, 4 (2001), 469–492.

[3] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2016. Enriching Word Vectors with Subword Information. *arXiv preprint arXiv:1607.04606* (2016).

[4] Sorathan Chaturapruek, Tobias Dalberg, Marissa E. Thompson, Sonia Giebel, Monique H. Harrison, Ramesh Johari, Mitchell L. Stevens, and Rene F. Kizilcec. 2021. Studying Undergraduate Course Consideration at Scale. *Contemporary Educational Psychology* 7 (2021).

[5] Shruthi Chockkalingam, Run Yu, and Zachary A. Pardos. 2021. Which One's More Work? Predicting Effective Credit Hours between Courses *(LAK21)*. Association for Computing Machinery, New York, NY, USA, 599–605. https://doi.org/10.1145/3448139.3448204

[6] Hendrik Heuer and Andreas Breiter. 2018. Student Success Prediction and the Trade-Off between Big Data and Data Minimization. *DeLFI* (2018).

[7] Mohammed E. Ibrahim, Yanyan Yang, David L. Ndzi, Guangguang Yang, and Murtadha Al-Maliki. 2019. Ontology-Based Personalized Course Recommendation Framework. *IEEE Access* 7 (2019), 5180–5199. https://doi.org/10.1109/ACCESS.2018.2889635

[8] Weijie Jiang and Zachary A Pardos. 2020. Evaluating sources of course information and models of representation on a variety of institutional prediction tasks. In *Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020)*. 115–125.

[9] Weijie Jiang and Zachary A Pardos. 2021. Towards equity and algorithmic fairness in student grade prediction. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 608–617.

[10] Naresh K. Malhotra. 1982. Information load and consumer decision making. *Journal of Consumer Research* 8, 4 (1982), 419–430.

[11] Sara Morsy and George Karypis. 2019. Will This Course Increase or Decrease Your GPA? Towards Grade-Aware Course Recommendation. *Journal of Educational Data Mining* 11, 2 (2019), 20–46.

[12] Alice Peasgood Niall Sclater and Joel Mullan. 2016. Learning analytics in higher education. A review of UK and international practice. *Jisc* (2016).

[13] Aditya Parameswaran, Petros Venetis, and Hector Garcia-Molina. 2011. Recommendation Systems with Complex Constraints: A Course Recommendation Perspective. *ACM Trans. Inf. Syst.* 29, 4, Article 20 (dec 2011), 33 pages. https://doi.org/10.1145/2037661.2037665

[14] Zachary A. Pardos, Hung Chau, and Haocheng Zhao. 2019. Data-Assistive Course-to-Course Articulation Using Machine Translation. In *Proceedings of the Sixth (2019) ACM Conference on Learning @ Scale* (Chicago, IL, USA) *(L@S '19)*. Association for Computing Machinery, New York, NY, USA, Article 22, 10 pages. https://doi.org/10.1145/3330430.3333622

[15] Zachary A. Pardos, Zihao Fan, and Weijie Jiang. 2019. Connectionist Recommendation in the Wild: On the Utility and Scrutability of Neural Networks for Personalized Course Guidance. *User Modeling and User-Adapted Interaction* 29, 2 (apr 2019), 487–525. https://doi.org/10.1007/s11257-019-09218-7

[16] Zachary A. Pardos and Weijie Jiang. 2020. Designing for Serendipity in a University Course Recommendation System. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (Frankfurt, Germany) *(LAK '20)*. Association for Computing Machinery, New York, NY, USA, 350–359. https://doi.org/10.1145/3375462.3375524

[17] Agoritsa Polyzou and George Karypis. 2019. Feature Extraction for Next-Term Prediction of Poor Student Performance. *IEEE Transactions on Learning Technologies* 12, 2 (2019), 237–248. https://doi.org/10.1109/TLT.2019.2913358

[18] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Conference on Empirical Methods in Natural Language Processing*.

[19] Josipa Roksa and Tania Levey. 2010. What can you do with that degree? College major and occupational status of college graduates over time. *Social Forces* 89, 2 (2010), 389–415.

[20] Erzhuo Shao, Shiyuan Guo, and Zachary A. Pardos. 2021. Degree Planning with PLAN-BERT: Multi-Semester Recommendation Using Future Courses of Interest. *Proceedings of the AAAI Conference on Artificial Intelligence* 35, 17 (May 2021), 14920–14929. https://doi.org/10.1609/aaai.v35i17.17751

[21] Shanna Smith Jaggars Thomas Bailey and Davis Jenkins. 2015. Redesigning America's community colleges. *Harvard University Press* (2015).

[22] María Cora Urdaneta-Ponte, Amaia Mendez-Zorrilla, and Ibon Oleagordia-Ruiz. 2021. Recommendation Systems for Education: Systematic Review. *Electronics* 10, 14 (2021). https://doi.org/10.3390/electronics10141611

[23] Nhi N.Y. Vo, Quang T. Vu, Nam H. Vu, Tu A. Vu, Bang D. Mach, and Guandong Xu. 2022. Domain-specific NLP system to support learning path and curriculum design at tech universities. *Computers and Education: Artificial Intelligence* (2022). https://www.sciencedirect.com/science/article/pii/S2666920X21000369

[24] Renzhe Yu, Qiujie Li, Christian Fischer, Shayan Doroudi, and Di Xu. 2020. Towards Accurate and Fair Prediction of College Success: Evaluating Different Sources of Student Data. In *Educational Data Mining*.

[25] Brian Hu Zhang, Blake Lemoine, and Margaret Mitchell. 2018. Mitigating Unwanted Biases with Adversarial Learning. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (New Orleans, LA, USA) *(AIES '18)*. Association for Computing Machinery, New York, NY, USA, 335–340. https://doi.org/10.1145/3278721.3278779