

Diagnosing University Student Subject Proficiency and Predicting Degree Completion in Vector Space

Yuetian Luo

UW-Madison, USA

yluo86@wisc.edu

Zachary A. Pardos

UC Berkeley, USA

zp@berkeley.edu

Abstract

We investigate the issues of undergraduate on-time graduation with respect to subject proficiencies through the lens of representation learning, training a student vector embeddings from a dataset of 8 years of course enrollments. We compare the per-semester student representations of a cohort of undergraduate Integrative Biology majors to those of graduated students in subject areas involved in their degree requirements. The result is an embedding rich in information about the relationships between majors and pathways taken by students which encoded enough information to improve prediction accuracy of on-time graduation to 95%, up from a baseline of 87.3%. Challenges to preparation of the data for student vectorization and sourcing of validation sets for optimization are discussed.

Introduction

We propose a novel approach to representing students' course paths by applying word vector models to course enrollment sequences. Students' per-semester vectors, which represent their course history, are then used as input to train a predictive models of on-time graduation. We explore if the semantics of the space allow for diagnosing student gaps in proficiencies through their vector similarity to representations of subject degree requirements.

Related Work

Keeping time to degree near the four-year expected norm is a priority of colleges and universities as extending beyond that time stresses the finances of both the student and financial aid departments and can reflect poorly on the institution's educational mission. Various reports, including from Higher Education Research Institution at UCLA (2004), show that this problem is widespread, with only 38.9% of undergraduate student across the country com-

pleting a degree after four years and 56.4% after five. The problem is worse among public institutions (37.1%) compared with private (64%). A more recent national survey by the National Student Clearinghouse Research Center (2016) shows the situation has not changed for four-year public institutions, with 37.5% students graduating within the intended four-year period.

College completion has long since been a focus of researchers. Much of the academic groundwork around student dropout was done in the 1970s-1980s. Tinto (1975) distinguished different types of dropout, such as dropout resulting from academic failure or voluntary withdrawal and permanent or temporary dropout and a theoretical model to explain the process. Bean (1980) developed a causal model to investigate the determinants of student attrition in institutions of higher education. They found that institutional commitment (loyalty to the organization) and routinization were two of the most significant factors that affected dropout. Recently, with readily available computing resources and the ubiquity of data, the student dropout problem has been revived, analytically. Dekker et al. (2009) use a host of machine learning approaches to predict student dropout among a group of 648 students in the Electrical Engineering department at the Eindhoven University of Technology using the first semester's grades. They have shown that a simple and intuitive classifier could give a result with accuracies between 75%-80%. In Bayer et al. (2012)'s work, they have shown a significant increase of prediction accuracy of drop out by taking students' social behavior into consideration. Aulck et al. (2016) modeled student drop out using the largest known dataset on higher education attrition, which tracks over 32,500 students' demographics and transcript records. They found that GPA in Math, English, Chemistry and Psychology classes were among the strongest individual predictors of student retention. If those factors are known, a hypothesis has been that an early warning system could be instituted to deploy appropriate interventions before the student drops out. Jayaprakash et al. (2014) trialed such an

effort called the Open Academic Analytics Initiative (OAAI). They found that relatively simple intervention strategies designed to alert students early in a course can positively impact student learning outcomes, but the intervention could also have unintended consequences such as triggering students to withdraw from the course, often early in the semester, as a means to avoid academic and financial penalties.

Dropout research has similarly been revitalized in application to the informal post-secondary context of Massive Open Online Course (MOOCs), where drop out prediction can utilize the clickstream data collected from learner interactions with the course. Whitehill et al. (2017) computed the accuracy of a standard logistic model of drop out prediction using features engineered from such clickstream data. Yang et al. (2013) explored the effect of students' behavior and social positioning on their drop out and found that forum posting behavior and social centrality measures had significant influence on drop out.

Drop out and pathway modeling (Lin, 2009; Imbrie et al., 2008) are both topics which relate to the study of on-time graduation. Research directly on time to degree (TTD) completion has not been investigated as comprehensively. Herzog (2006) did investigate the issue and used neural networks, decision trees, and logistic regression to predict time to degree with seventy-nine variables and achieved an accuracy around 83% for completion times three years or less and 93% for completion times six years or more.

Representation learning, our chosen model of analysis, is an emerging mainstay in computational linguistics. One of the most popular methods is the skip-gram (Mikolov et al., 2013a), coined "word2vec" after the name of its open sourced package (Mikolov et al., 2013b). Recently, the concept of distributed representations has been expanded beyond word representation to many other fields such as graph structure (Perozzi et al., 2014), e-commerce (Grbovic et al., 2015), fashion, and online tutoring systems (Pardos & Dadu, 2017). In our research, we continue the work of representing courses as vectors (Pardos & Nam, 2017) to, instead, construct student representation and depict their learning path. Using student continuous vector representations as features, we then predict time to degree using leave-one-out cross-validation.

Dataset

We used a dataset from the University of California at Berkley (UCB) which contained anonymized student course enrollments information from the Fall of 2008 through Fall 2015. The dataset consisted of per-semester course enrollment information for 99,971 undergraduates, 38,466 graduates and 22,814 visiting summer exchange students with a total of 3.6M course enrollment records. A course enrollment meant that the student was still enrolled

in the course at the end of the semester. Some records of student enrollment data are shown in Table 1. In Table 2 we show the on-time¹ graduation statistics of all freshman undergraduates entering school in Fall 2008, 2009, and 2010.

ANON ID	Semester Year	Undergraduate/ Graduate	Dept.	Course Number	Major
x028224	Spring 2014	Undergrad	Law	178	Law
x028224	Summer 2014	Undergrad	Law	165	Law
x028224	Fall 2014	Undergrad	Math	140	Math
x028224	Fall 2014	Undergrad	Math	121	Math

Table1. Example of source dataset of student enrollments.

Year	On-time graduates	Overtime graduates
2008	3,226	758
2009	3,225	711
2010	3,227	529

Table 2. Number of on-time and overtime graduation for Fall 2008-2010 freshman undergraduates.

In Table 3, we give the top five majors with the highest number of overtime undergraduates for this set of students and the within-major percentage this represented.

Major	Overtime graduate number
Integrative Biology	87 (12%)
Interdisciplinary Studies	61 (14%)
American Studies	55 (11%)
Political Economy	49 (9%)
Sociology	44 (9%)

Table 3. Top 5 majors by overtime graduations.

From Table 3, we can see that Integrative Biology had the largest number of students who graduated overtime. In this paper, these students serve as our case study for overtime analysis. The detailed graduation status by year of Integrative Biology majors is shown in Table 4.

Integrative biology	2008	2009	2010
On-time graduate (≤ 4 years)	277	258	250
Overtime graduate (> 4 years)	36	27	24

Table 4. Number of on-time and overtime graduation undergraduates from 2008 to 2010.

Methods

Skip-gram Model

In natural language processing (NLP), the training objective of the skip-gram model is to find word representations

¹ On-time graduation refers to attaining a Bachelor degree within four years. Overtime refers to attaining the degree in more than four years.

that are predictive of the surrounding words in a sequence of natural language. The sequence of words can be symbolized as $X_1, X_2, X_3, \dots, X_T$, with the objective of the model being to minimize the average log loss:

$$\frac{1}{T} \sum_{t=1}^T \sum_{-c \leq j \leq c, j \neq 0} \log p(X_{t+j} | X_t)$$

where c is the size of the training context, or window size (a hyper parameter). The probability of a context word, X_{t+j} , is determined by a softmax function:

$$p(X_o | X_I) = \frac{\exp(v_{X_o}^T v_{X_I})}{\sum_{j=1}^W \exp(v_{X_j}^T v_{X_I})}$$

where X_o and X_I are output and input words; v_X and v_X' are “input” and “output” vector representation of X , and W is the number of words in the vocabulary. The product of the input word vector multiplied by the context word vectors (learned via backpropagation) creates the activations which are normalized into probabilities via the softmax function.

Constructing Student Representations

As shown in Pardos & Nam (2017), it is straightforward to construct “sentences” of courses for use with word2vec by tokenizing them based on subject_course# and sequencing them based on the time a student enrolled, randomizing the tokens of courses taken by a student within a single semester. In this pre-processing, one sentence is constructed per student and then presented to word2vec for training. This approach learned vector representations of courses. The research question at hand is how to apply this paradigm to represent students. Using RNNs would be a reasonable approach, as they have been used in education contexts to predict the correctness of future exercises in a tutor (Piech et al., 2015) and the next page which will be visited in a MOOC (Tang, Peterson, & Pardos, 2017). Given the efficient computational implementation of word2vec and the scrutable vector space it creates which allows for arithmetic and scalar manipulation of the representations, we chose to continue to adopt this model to the task of student vector representation. The key to the construction is to create “sentences” with student IDs as the tokens, instead of course IDs. To achieve this, we first group students by each course concatenated with the semester and year it was offered in, along with the grade received in the course. Table 5 shows an example of this formatting.

Data for CS100 Fall'10	Skip-gram training data format (each row is “sentence”)
stu1 A, stu2 B, stu3 C,	stu4 stu1
stu4 A, stu5 C, stu6 B,	stu2 stu6 stu7
stu7 B, stu8 C	stu5 stu8 stu3

Table 5. Example of student token formatting for training with word2vec.

Finally, we randomize the list of student IDs in each group and then concatenate it with other ID lists in the same course with same grade by chronological order, such as Fall 2010 CS100’s student ID list which is concatenated with Fall 2011 CS100’s list and so on. The end result is that for each grade on a particular course, we get one “sentence.” The source of signal in this formatting is three fold. First, it has chronological structure such that students who take the same courses in the same semester are closer to one another, which can be learned in the representation. Second, students who are in the same major are more likely to share courses with one another. It is this network of shared courses that will produce the embedding of students based on the collection of courses they took and when they took them. Lastly, by grouping courses by grade, we allow for paths to be described in the dimension of student success in those courses, and not just courses taken. Word2vec can now be run on these course_semester_grade groupings to produce the representations of students given their progress, thus far, in their undergraduate career.

Constructing Intermediary Semester Representations

The student representation method previously explained can represent a student’s entire course history, but in this work, we want to follow several cohorts of Integrative Biology students one semester at a time, evaluating their proficiencies and estimating their chances of on-time graduation according to a shared vector space. To do this, we need representations of those students at every semester, not just a compilation of their entire course history at their last semester. In this section, we introduce how we allowed word2vec to represent students at each semester.

We consider an on-time student, “A,” and an overtime student, “B,” to illustrate the construction process. First, in order to get student A and B’s representation after the first semester, we create two copies named A_ontime_sem1 and $B_overtime_sem1$ in the original data. A_ontime_sem1 has only student A’s first semester’s courses and $B_overtime_sem1$ only has student B’s first semester’s courses. After using the technique to create the modified word2vec training data, the learned vector of A_ontime_sem1 and $B_overtime_sem1$ would represent student A and B’s vector after the first semester. We continue with this duplication process for each semester, which would end at the eighth semester for on-time student A and would continue until the graduating semester for student B.

Model Training

For word embeddings, model selection based on hyper parameters is done according to the model’s score on semantic and syntactic sets of word relationship validation sets. Pardos & Nam (2017) do model selection in the

course context by minimizing the nearest neighbor ranks of pairs of cross-listed courses. For optimizing student representations, as is the focus of this paper, it is not clear where a reasonable validation might come from given our dataset, which did not contain any student demographic information. In order to address this problem, we use an intuition from the course vector context and optimize for student vector groupings by their declared major.

We first develop an intuition for the correlation between course vector clustering by subject and the performance of a course embedding on an acceptable validation set (cross-listings). We ran k-means clustering for course vectors using cosine as the distance metric, and set the cluster number equal to the number of course subjects² in the dataset. We then optimized assigning each cluster to one subject by using the Hungarian (Kuhn et al., 1955) method. Specifically, the gain of assignment of one cluster to the subject of ‘Mathematics’ is the number of students who are in the Math major in that particular cluster. So after iterating for all clusters and subjects, we can get the gain matrix for the Hungarian method. After we find the best assignment, we could then calculate its assignment accuracy. Say there were N different subjects and N clusters of courses. For each course in each cluster, if it belongs to the assigned subject of that cluster, the number of correct answers, Y, is incremented by one. After iterating all courses in all clusters, the course assignment accuracy equals to Y divided by the total number of courses.

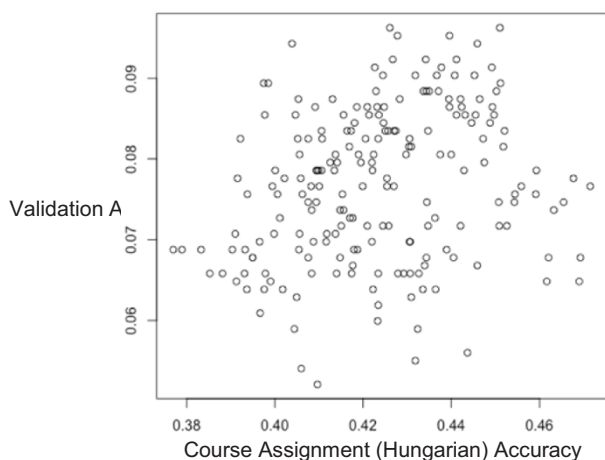


Figure 1. Scatter plot of course assignment accuracy and cross-list validation set accuracy

We found that there is a positive relationship between course assignment accuracy and cross-listed accuracy (Figure 1). Running a regression, we calculated a coefficient of 0.499 and p-value of 0.00062, allowing us to reject

² A ‘Subject’ is the smallest category of academic unit in the University

the null hypothesis and suppose there is a positive relationship between course assignment accuracy and the cross-listed validation set. We then make a jump in assuming that if maximizing course clustering by subject in model selection resulted in a quality embedding, then, given no better alternatives, selecting a model by maximizing student clustering by major may similarly result in a quality student embedding. We proceeded to train 130 models, varying the vector length between 48 and 192 and the window size between 10 and 130 and selected the model with highest accuracy according to the Hungarian method.

Visualization and Proficiency Estimation

In our research, we wish to show the richness of information contained in the student representation, visually. To do so, we chose t-SNE for dimensionality reduction (Maaten, Hinton; 2008). Default parameters of perplexity 30, PCA to initial dim of 50, and theta 0.50 were used.

In the previous sections, we introduced methods to construct the students’ enrollment paths in order to generate per semester vector representations of them. The purpose in doing this was to be able to relate the vectors of our Integrative Biology cohorts, at each semester, to notions of proficiency in the vector space. In this section, we talk about the process of creating these notions of proficiency from average student major vectors and using cosine similarity to evaluate a student’s semester vector with respect to those averages. We use the full enrollment sequence of freshman entering in 2008-2009 to generate averages of those students by major and use the per semester vectors of Integrative Biology freshman entering in Fall 2009 to relate to those average major vectors. The assumption is that the average graduated major was at an advanced stage of proficiency in their major area at graduation, at least compared to non-majors. Integrative Bio (IB) majors who are “more like” the average graduated student in a particular major, are perhaps more likely to have that major’s proficiencies than students who are “unlike” them. Using the IB students entering in Fall 2008 and graduating on-time, we created an average of their student major vectors as of Spring 2012, named **Ave_On**. This would be used to compare the progress, by semester, of IB students entering in Fall 2009 to graduated IB students. The Integrative Biology major has courses in Physics, Chemistry and Math listed³ as required for the major, thus we calculated the average vectors **Ave_Phy**, **Ave_Che**, **Ave_Math** of on-time graduated students in those majors who entered in Fall 2008. The semester vectors for 2008 IB freshman could then be compared to those subjects and to the vector

³ <http://guide.berkeley.edu/undergraduate/degree-programs/integrative-biology/#majorrequirementstext>

of IB students in the previous year's cohort who graduated. We note that this IB to IB comparison would not be possible in real-time, since 2008 students' graduated vectors would not be available for comparison in 2009. We suggest that if more prior year's data were available, an IB vector of graduated students could have been produced based on students who graduated in 2009, albeit with more differences in pathway, due to changing course offerings. The IB to IB analysis will nevertheless serve as a proof-of-concept, for comparing this cohort to successful students in a past cohort.

Prediction

With IB students' semester vectors in hand, we attempted to predict whether the student, at each semester, could graduate on-time. We used two algorithms to conduct the prediction; logistic regression and neural networks. The candidate features serving as input were the student's semester vector representations and the cosine similarity, dot product, and Euclidian distance to each of the four average major vectors (Integrative Biology, Physics, Chemistry, Math). The label was a binary variable representing whether the student graduated on-time. The evaluation was conducted using leave-one-out cross-validation (to maximize training set size) at the student level, such that when a student was left out of training, all of his or her semester vec-

tors would be the target of prediction in that phase of the cross-validation. For the neural network, we used a single hidden layer with 200 nodes, sigmoid activation function, and Adagrad optimizer. The loss function was binary cross-entropy with a max epoch of 10.

Results

Student Embedding Visualization

The best trained student embedding, according to the Hungarian method, had a window size of 90, vector dimension of 96 (with 30 iterations and min student count of 3). The t-SNE projection of this embedding can be seen in Figure 2 where each circle represents one student, colored by declared major. We can first observe that students are generally clustered together by major. This was the explicit objective of the Hungarian method optimization in the higher dimensional space, but it is also a logical expectation regardless of optimization, as it was seen that courses cluster by subject without explicitly optimizing for that. The largest cluster of students is the orange in the lower left which represents those who have not yet declared a major. In the upper left a cluster of natural sciences and engineering can be seen, consisting of Chemical Engineering, Mechanical Engineering, Electrical Engineer and Computer Science (EECS), and Bioengineering. We can also observe that

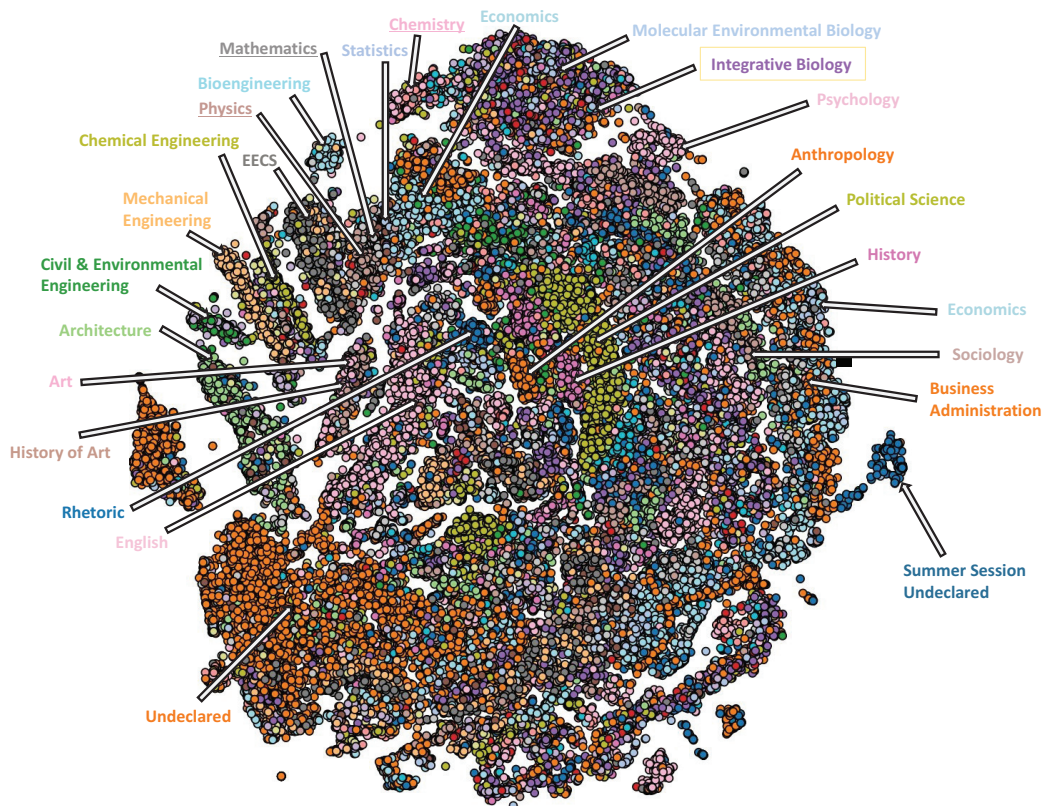


Figure 2. t-SNE projection of the embedding of all students in the dataset. Major labels are colored by their respective cluster of data points. Majors relating to Integrative Biology degree requirements are underlined.

many Applied Mathematics and Computer Science students are near one another with the Engineering students. Another interesting observation is that we can find Economics students at two modal locations in the space. The first location is on the right side, near the Social Science student group, and on the left side, near the Mathematics and Engineering group. This is reasonable given the two major foci in undergraduate Economics. Lastly, at the very right side, we can find an isolated group of students who are summer visitors. We present these observations, which only scratch the surface of what can be found, as sanity checks on the semantics of the plot⁴.

Subject Proficiency Analysis

In this section, we show the results of Integrative Biology student's cosine similarity with graduated student averages from Physics, Math, Chemistry, and Integrative Biology itself (from the prior year's eventual on-time graduates). In Figure 3, the cosine similarity of each IB student at each semester is calculated with respect to the subject averages with eventual on-time students colored green, and overtime students, red.

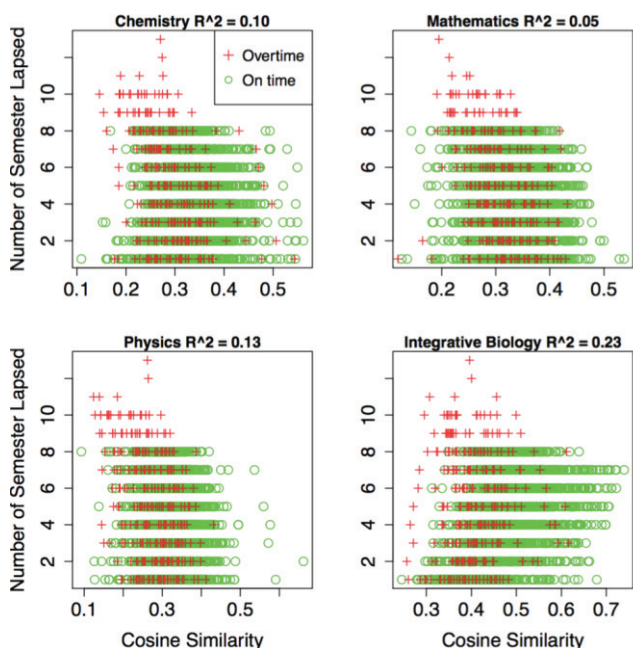


Figure 3. On-time and overtime student-semester cosine similarity with averaged subject vectors relevant to the IB major.

From Figure 3, we can see that for Chemistry, Physics, and Integrative Biology, the cosine similarities of on-time students are slightly higher than overtime students. For the Math subject, it seems there is no difference. Perhaps because Mathematics is less crucial a subject in IB or because

Mathematics students are so disjoint from IB that their vector shares few features with the IB students' vectors. The general trend of on-time students having higher cosine similarity to a comparison group continues in Figure 4, where the comparison is to the average of the IB students' own vectors. This trend perhaps works out due to the fact that most IB students graduate on-time.

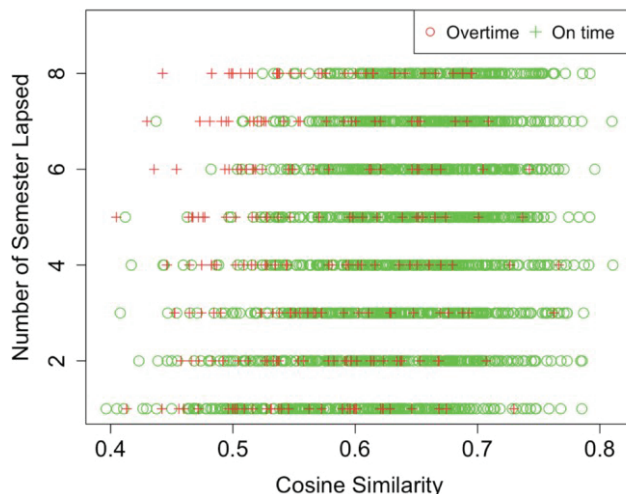


Figure 4. On time and overtime cosine similarity projection by semester on themselves (relative within cohort proficiency).

On Time Graduation Prediction

After creating representations for each student-semester of our 2009 IB cohort, the number of on-time samples was 1,972 and the number of overtime samples was 288 (87.2% on-time). The total number of IB students was 285. After training the neural networks with a variety of feature combinations, the average accuracy results of predicting on-time graduation by semester was evaluated and is shown in Table 6 and Figure 5.

Method	Input				Ave. Acc.
	Vector	Euclidian	Cosine similarity	Dot projection	
Logistic	X				0.880
		X			0.873
			X		0.873
				X	0.873
	X	X	X	X	0.895
Neural Network 1 layer 200 nodes	X				0.955
		X			0.873
			X		0.874
				X	0.873
	X	X	X	X	0.956

Table 6. On-time graduation prediction accuracy using two models with five input feature combinations.

⁴ The d3 plot software: <https://github.com/CAHLR/d3-scatterplot>

From Table 6, we find that single layer neural network model is better than the logistic regression for extracting information from the representation alone (88% vs 95.5%). The vector representations alone performed better than the individual cosine similarity, dot product, and Euclidian distance statistics alone. Those statistics added a minor bump in performance when added to the representations for both logistic (88% vs 89.5%) and the neural network (95.5% vs 95.6%). In Figure 5, we can see that from semester one to semester eight, the prediction performance of the neural network model consistently outperformed the logistic model, particularly in the overtime semesters where the logistic primarily predicts the majority class. A logistic regression with vectors as input is a dot product of the input vector with a learned positive class vector (the weight coefficients). The positive performance of the neural network classifier suggests that there is a distributed representation of overtime graduation in the embedding but that it is not encoded locally as simply magnitude or direction (angle), thus posing a challenge to the magnitude and direction (dot-product) based logistic classification.

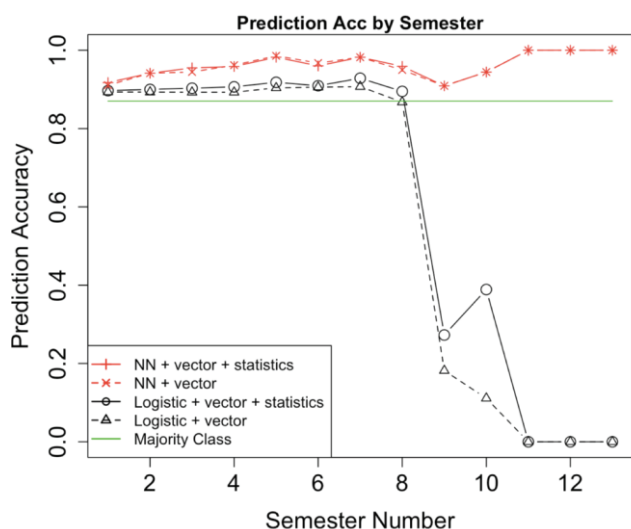


Figure 5. Prediction Accuracy of Best Logistic and Neural Network Model by Semester.

Conclusion

This research introduced an exploratory methodology to study how a student's state, representing their course history, correlates to notions of proficiency, strictly keeping to manipulations in the vector space. In all but one of the 32 on-time subject-semesters in the four proficiency plots, the student with the highest cosine value was an on-time student; however, in 15 of the 32, the student with the lowest cosine was not an overtime student. This suggests that having proficiencies is an indicator of on-time graduation but the absence of proficiencies is not necessarily an indicator to the contrary. Correlation of the student's state to Mathematics was the lowest among the three required subject

areas. It is an open question if this is due to the Mathematics vector being too far from Integrative Biology students to be compared to one another, or if Mathematics proficiency is simply not well correlated with success in on-time graduation in Integrative Biology. The second contribution of the work was to introduce a method for vector space visualization of student enrollment pathways. We scratched the surface of analyzing this vector space visualization which depicted logical major adjacencies at the University.

Future work can focus both on unearthing additional findings from the visualization and potentially surfacing proficiency like dashboard analytics for advisors so they may be informed of the knowledge gaps of their cohort and can focus their efforts on correcting their students' path. Additional future work includes consideration of all majors, their proficiencies, and time to degree prediction.

Acknowledgements

We thank the Office of the Registrar, Enterprise Data Warehouse, Education Technology Services, and Office of Planning and Analysis at UC Berkeley for their data access and documentation support. This work was funded by a grant from the National Science Foundation (Award #1547055).

References

- Clancey, W. J. 1984. Classification Problem Solving. In *Proceedings of the Fourth National Conference on Artificial Intelligence*, 49-54. Menlo Park, Calif. AAAI Press.
- Bayer, J., Bydzovská, H., Géryk, J., Obšivac, T., & Popelinský, L. 2012 Predicting drop-out from social behaviour of students. In *Proc. of the 5th International Conference on Educational Data Mining*. pp. 103-109.
- Bean, J. P. 1980. Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in higher education*, 12(2), 155-187.
- DeAngelo, L., Franke, R., Hurtado, S., Pryor, J. H., & Tran, S. 2011. Completing college: Assessing graduation rates at four-year institutions. *Los Angeles: Higher Education Research Institute, UCLA*.
- Dekker, G.W., Pechenizkiy, M., Vleeshouwers, J.M. 2009. Predicting Students Drop Out: A Case Study. In *International Conference on Educational Data Mining*. Cordoba, Spain, pp. 41-50.
- Grbovic, M., Radosavljevic, V., Djuric, N., Bhamidipati, N., Savla, J., Bhagwan, V., & Sharp, D. 2015. E-commerce in your inbox: Product recommendations at scale. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1809-1818). ACM.
- Herzog, S. 2006. Estimating student retention and degree-completion time: Decision trees and neural networks vis-à-vis regression. *New directions for institutional research*, (131), pp. 17-33.
- Imbrie, P. K., Lin, J., & Malyscheff, A. 2008. Artificial intelligence methods to forecast engineering students' retention based

- on cognitive and non-cognitive factors. In *Proceeding of the American Society for Engineering Education Annual Conference & Exposition*.
- Jayaprakash, S. M., Moody, E. W., Lauria, E. J., Regan, J. R., & Baron, J. D. 2014. Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics*, 1(1), 6-47.
- Kuhn, H. W. 1955. The Hungarian method for the assignment problem. *Naval Research Logistics (NRL)*, 2(1-2), 83-97.
- L Aulck, N Velagapudi, J Blumenstock, J West 2016. Predicting Student Dropout in Higher Education. *International Conference on Machine Learning (ICML) Workshop on Data for Good*.
- Lin, J. J., Imbrie, P. K., & Reid, K. J. 2009. Student retention modelling: An evaluation of different methods and their impact on prediction results. *Research in Engineering Education Symposium*, 1-6.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. 2013a. Efficient estimation of word representations in vector space. *arXiv preprint:1301.3781*.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. 2013b. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119).
- Pardos, Z. A., & Nam, A. J. H. 2017a. The iSchool of Information and its Relationship to Computer Science at UC Berkeley. In *Proceedings of the iConference* (pp.309-316).
- Pardos, Z.A., & Dadu, A. 2017b. Imputing KCs with Representations of Problem Content and Context. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*. Bratislava, Slovakia. ACM. Pages 148-155.
- Perozzi, B., Al-Rfou, R., & Skiena, S. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 701-710). ACM.
- Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. 2015. Deep knowledge tracing. In *Advances in Neural Information Processing Systems* (pp. 505-513).
- Shapiro, D., Dundar, A., Wakhungu, P. K., Yuan, X., Nathan, A., & Hwang, Y. 2016. Time to degree: A national view of the time enrolled and elapsed for associate and bachelor's degree earners. *Signature Report*, (11).
- Tang, S., Peterson, J., and Pardos, Z. 2017. Predictive Modelling of Student Behaviour Using Granular Large-Scale Action Data. In Lang, C., Siemens, G., Wise, A. F., and Gaevic, D., editors, *The Handbook of Learning Analytics*, pages 223–233. Society for Learning Analytics Research (SoLAR), Alberta, Canada, 1st edition.
- Tinto, V.1975. Dropout from higher education: A theoretical synthesis of recent research. *Review of educational research*, 45(1), 89-125.
- Whitehill, J., Mohan, K., Seaton, D., Rosen, Y., & Tingley, D. 2017. Delving Deeper into MOOC Student Dropout Prediction. *arXiv preprint:1702.06404*.
- Yang, D., Sinha, T., Adamson, D., & Rosé, C. P. 2013. Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *Proceedings of the NIPS Data-driven education workshop* (Vol. 11, p. 14).