Spring 2015

Modeling the Progress and Retention of International Students Using Markov Chains

Lucas Gagne
University of Akron Main Campus, lmg64@zips.uakron.edu

Modeling the Progress and Retention of International Students using Markov Chains

Lucas Gagne

Department of Mathematics

Honors Research Project

Submitted to

The Honors College

Approved:

TSNorfolk Date 12/3/14
Honors Project Sponsor (signed)

T. S. NORFOLK
Honors Project Sponsor (printed)

Gerald W Young Date 12/2/14
Reader (signed)

Gerald W Young
Reader (printed)

Curtis B. Clemons Date 12/3/14
Reader (signed)

Curtis B. Clemons
Reader (printed)

Accepted:

TSNorfolk Date 12/3/14
Department Head (signed)

T. S. NORFOLK
Department Head (printed)

_________________ Date ________
Honors Faculty Advisor (signed)

____
Honors Faculty Advisor (printed)

_________________ Date ________
Dean, Honors College

Modeling the Progress and Retention of International Students using Markov Chains

An Honors Research Project

Lucas Gagne

December, 2014

Abstract

International students are a small and diverse student population present in any sizable American university. One of the greatest obstacles in their path is the acquisition of the English language. English for Academic Purposes (EAP) programs, such as the English Language Institute (ELI) at the University of Akron, attempt to address this problem. By studying how this student population progresses in their academic studies, EAP programs and their associated universities can make well-informed decisions on how best to serve their English Language Learners. One way to study International students is through the use of a Markov model based on university data. The statistically calibrated model is helpful in that it can reveal at what stage of their studies students are struggling the most or succeeding the most. Our results show that ELI students are well prepared, since international students who are first time freshman seem to have a higher progression rate to the sophomore level than the standard STEM population of students. We also find that students who repeat their freshman year are more likely to repeat again than progress to the sophomore level.

Chapter 1

Introduction

The English Language Institute (ELI) at the University of Akron is an English for Academic Purposes program. Students from all over the world study in order to become proficient at English, and many students continue to pursue their studies at the University of Akron. International students are a valuable resource at the University of Akron, as they bring different ideas and cultures, as well as revenue and reputation to the University of Akron. In order to continue to attract, retain, and better serve these students, we need to understand how this student population progresses. In this paper we consider what happens to these students after they complete ELI and transition into becoming freshmen at the University of Akron.

While reviewing literature on this subject, we found three articles of particular interest. In [1] the students in a college-level English as a Second Language (ESL) course were surveyed, with the conclusion being that the perception of the effectiveness of the class was decidedly positive. 77.5 percent of students agreed or strongly agreed (gave ratings of 4 or 5 on the standard 5 point scale) that their class was effective. For our purposes, it is interesting to note that 90 percent of students reported that the course was helpful for other classes at the university, 77.5 percent reported the class helped them better communicate with other students, and 80 percent reported that the course was helpful for succeeding at the university. These statistics underscore the importance and value of ESL courses, as well as programs like ELI to international students who do not speak English natively.

Numerous factors determine an ESL international student's success [2]. These include self-efficacy, values, perceived difficulty, motivation, and self-regulation. The overall grade point average (GPA), English grade, and results from a questionnaire determining their alignment with the above factors were combined and compared. Obviously, the relationships between these sets of data are complicated, but it was shown to be that the key factors for an ESL international student's academic success are motivation, self-regulation, self-efficacy, and English proficiency. This further emphasizes the importance of English proficiency to international students here at the University of Akron.

Another factor that plays into the success of ESL students is the usage of content-linked programs [3]. ESL Students who are engaged in their English classes perform better in all of their classes, graduate faster, and have higher GPA's. If the University of Akron wants its ESL international students to succeed, it will place great importance on the ELI program. This leads to several natural questions.

Is the University of Akron's ELI program successful? Where are ELI students struggling and succeeding in their University of Akron careers? We intend to better understand the answers to these questions by modeling using a Markov chain.

You can use a Markov Chain to model the paths of student population through the class ranks [4]. A Markov chain is a mathematical structure which consists of a set of states and probabilities of transition between states. Two properties characterize Markov chains, as we use them throughout this paper. Firstly, we are considering discrete-time Markov chains. This means time will be handled not continuously, but in increments of single years. Secondly, the probability of transitioning depends only on the current state, and not on any previous states that were journeyed through. This is called the Markov property.

An example of a Markov Chain used in this fashion [5] created a model that was able to accurately predict STEM students' graduation rates in STEM and non-STEM fields (for those that changed their majors). The study reviewed 692 transcripts, and showed a 47 percent graduation rate (from the University of Akron within 6 years, not including co-op). Further, it was shown that the model shows the progression of students to be inelastic. This is to say that large adjustments to individual probabilities are required to incur small changes in the graduation rate. However, changing initial probabilities to model the recruitment of top students had an immediate impact on the graduation rate.

Chapter 2

Description of Data

The project began by getting student ID numbers from the International Office and pulling transcripts, all with IRB approval. The ID numbers that were pulled corresponded with any student that had spent time in ELI, tracking back 8 years. The data included students who began ELI 8 years ago and have since graduated, and students who have recently started ELI.

Once we had pulled the transcripts, we went through each transcript and coded it into a spreadsheet. We then looked at our population and threw out groups such as transfer students (about 50 students), students who only attended ELI (about 47), and students who had not started in ELI and were not relevant to the project (about 200). The most interesting of these 'castaways' is the group of students who only attended ELI. Did these students finish and leave to attend another University? Did they drop out of ELI? This question alone merits further study.

We were left with a population of about one hundred students that could be investigated, using the data from the spreadsheet. The small size of the population prevented any meaningful analysis beyond year two. After each year, students can be sent to an absorbing state, one from which no transition is possible. For example, a student who graduates cannot move to any other state within our model.

When coding, we first recorded the exact point from which we had pulled the information, so we could easily find the record in the future, if necessary. Next, we recorded the overall number of students we had coded. From there, we had nine columns, labeled from zero to seven and an eighth column labeled Final. These correspond to each year the student attended ELI or the University of Akron. A student who received no collegiate credit in one academic year was coded as ‘0’ for that year, indicating they were in ELI. As soon as a student took any class at the University of Akron, they were coded as a ‘1’ (a freshman). Once a student received 32 credits in an academic year, they were coded as ‘2’ (a sophomore), and so on. If at any point the student stopped taking classes, they were coded as ‘11’, corresponding to students who did not graduate. A student who graduated was coded as ‘10’. In this context, an academic year refers to a Summer, Fall, and Spring semester, but not necessarily in that order. An academic year for a student who took their initial classes in the spring would look like Spring, Summer, then Fall, for example.

Chapter 3

Overview of Model

Our example model [5] tracks the entire career of STEM students, from freshman year to graduation. The model initially did not have the states Repeat Freshman, Repeat Sophomore, Repeat Junior, and Repeat Senior, but after running this model it was found the model had to be adjusted due to inaccuracy of the model's predicted 6 year graduation rates. In this model (Figure 1), α, β, δ, γ, λ, and ε are the probabilities of students following that respective path. Generally speaking, α corresponds to progressing immediately from one class rank (strictly in terms of credits earned) to the next without repeating, β to leaving the university or surpassing the six year time limit, δ to repeating a class rank, γ to progressing to the next class rank after repeating the class rank before, λ to remaining in a class rank for at least two years, and ε to leaving the university after repeating a class rank. The subscript numbers correspond to the class rank that the student is leaving from. Thus each class rank has its own set of probabilities α, β, δ, γ, λ, and ε.

The model has absorbing states for not graduating from the University of Akron, graduating STEM, and graduating non-STEM. The distinction was especially relevant to their study because all of the students they tracked were declared STEM majors at some point in their collegiate career (by virtue of taking a Pre-Calculus or Calculus course), so it was interesting to observe if those students ended up in a STEM major or not.

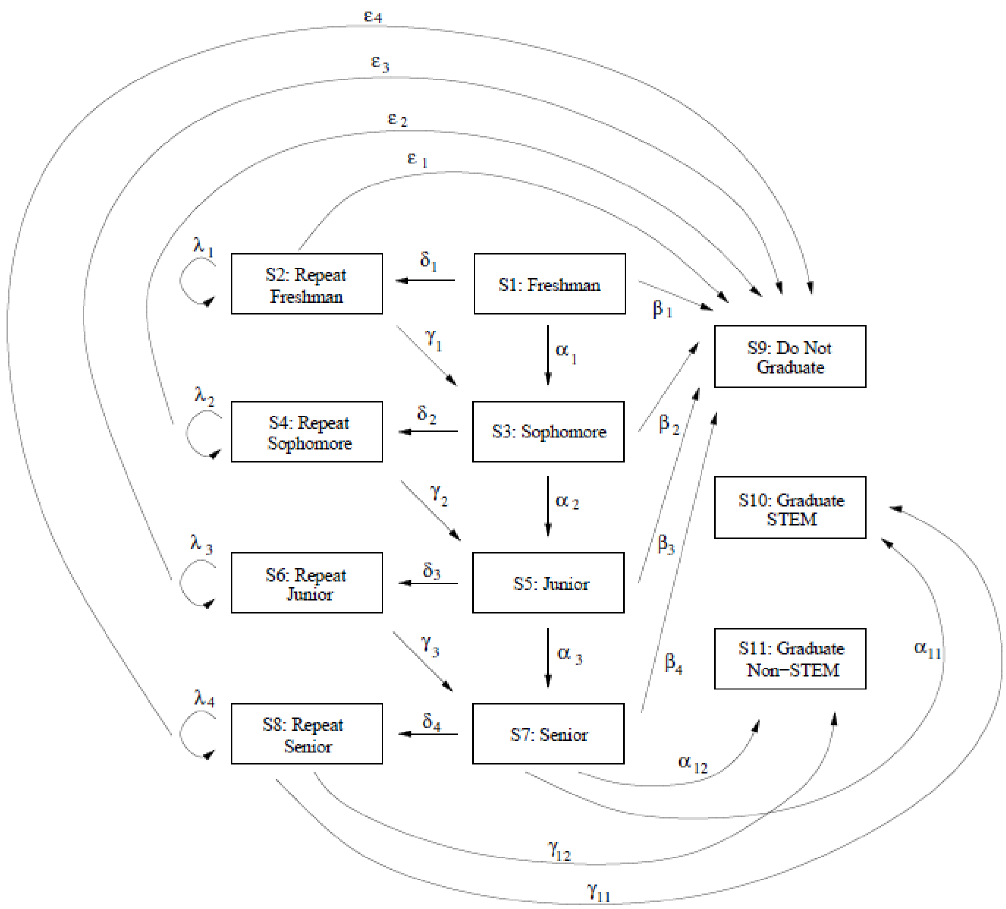

Figure 1. STEM model.

The study in [5] had a large amount of usable data and could create a model for the complete career (up to 6 years) (Figure 1) of an Akron student. The much smaller population of this project demanded we look at a smaller snapshot of a student’s career. Our model looks at a student who begins in ELI and can only model their progress until they reach the sophomore level. This model holds ‘not graduating’ and ‘sophomore’ as absorbing states. Further, our model only tracks the student once they have left ELI and begun taking classes at the University of Akron. It is important to note that only students who began their University of Akron career in ELI were considered in this study.

Chapter 4

Adapting the Model

If we consider only states S1, S2, S3, S9, and S10 (Figure 1), with some adjustment, we have a suitable model for our purposes. This is to say that this model can suit the smaller amount of data and consider the population of students who started their University of Akron undergraduate career in ELI (as opposed to STEM students). The smaller amount of data in this study (as explained in Chapter 3) precludes our study to freshmen only. Our model (Figure 2) includes a state for ELI, but this is simply to be clear that all students in our model began in ELI. We are only considering students who progressed into freshman status, so the transition probability would be trivial (1).

In this model α and γ correspond to the students who progress, β and ε to the students who leave the university, and δ and λ to the students who repeat some state in the model. We will view α, γ, β, ε, δ, and λ as probabilities.

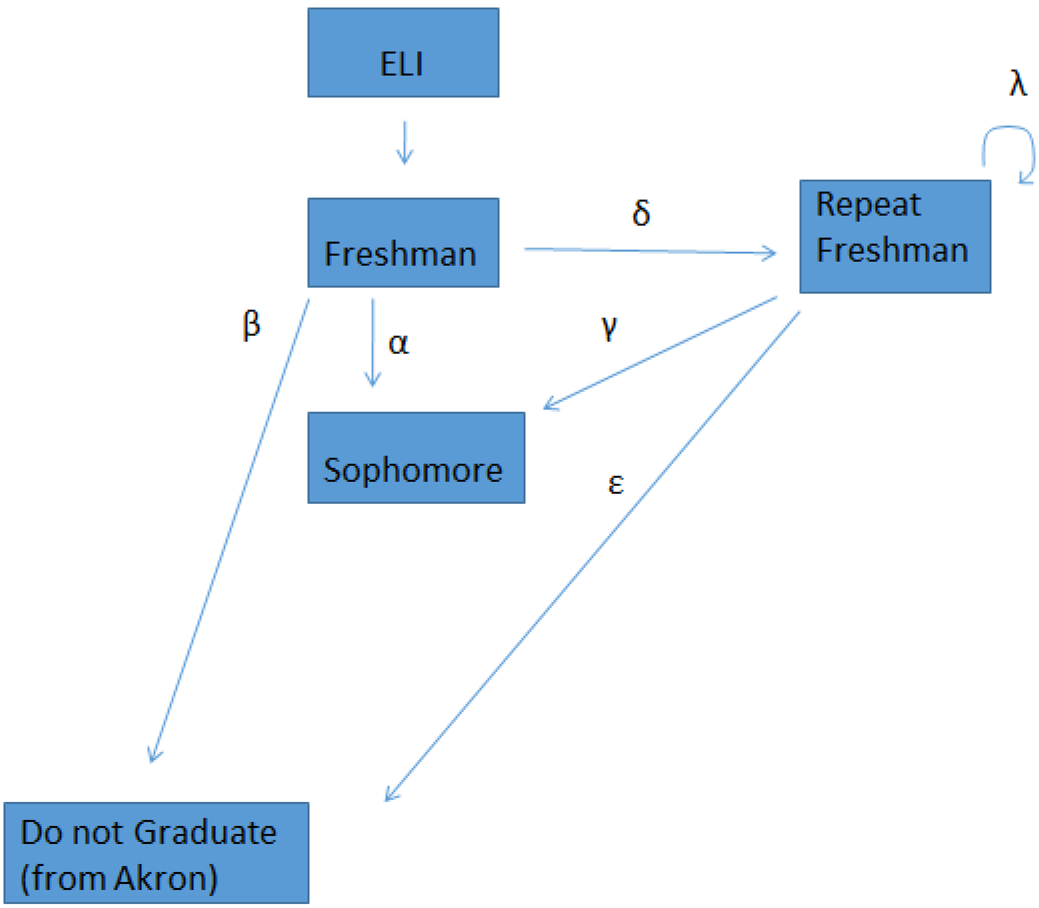

Figure 2. Adapted ELI model.

We define α as the probability that a student will progress from a first year freshman to sophomore year, bearing in mind that the terms freshman and sophomore are only indicative of the number of credits completed. β is the probability that a student will leave the university after being a first year freshman. δ refers to students who, after one academic year, do not have the necessary credits to be considered a sophomore.

It was found that the students who repeated their freshman year progressed at a rate distinctly different than first time freshman (α versus γ) [5]. It was necessary, in order to make an accurate model, to create a new state for these repeat freshman.

The transition of a repeating freshman who does not gain enough credits to be a sophomore is represented by λ. Students who, after 2 or more years of having only enough credits to be considered a freshman, progress to the sophomore level are then represented by γ. Finally, the probability that a repeat freshman will leave the university is labelled as ε.

Chapter 5

Analysis

Now that we have the model established, we can discuss the analysis of the data. To find our probabilities, we take our sample students (the data) and watch what happens as each academic year progresses. In this case it takes 4 academic years for all of our students to fall into an absorbing state. We draw our probabilities from what we observe from our sample. They are summed up in the following table. The reader should note from the previous section that coding 1 corresponds with freshman, 2 with sophomore, and 11 with not graduating. In addition, r1 refers to repeat freshman. All probabilities are rounded to the third decimal place.

Table 1. ELI model transition probabilities.

α	β	δ	γ	ε	λ
1 to 2	1 to 11	1 to r1	r1 to 2	r1 to 11	r1 to r1
.500	.180	.320	.381	.140	.479

We find that first-time freshman have a significantly higher chance of progressing to the sophomore level than do repeating freshman (Table 1). It is also interesting that repeating freshman leave at a lower rate, as well as repeat at a high rate. It is almost as if, for repeating freshman, the probabilities for progressing and repeating were swapped when compared to first time freshman.

Out of the 100 in our sample, 8 graduated from the University of Akron, 56 left the university, and 36 are still attending the university (or were at the time the data was pulled). Ignoring the continuing students, this gives us a 12.5% graduation rate. This number should not be taken seriously, as it is unlikely that the 36 still in attendance will follow that trend. Students who leave, usually do so early in their career, so those still continuing should have a much higher graduation rate. Of those who graduated, only 1 repeated their freshman year. However, 4 students took five years to graduate, while the other 4 students took six years to graduate.

Chapter 6

Conclusion

The comparison of resulting probabilities [5] are included below (Table 2).

Table 2. Comparison of ELI and STEM transition probabilities.

	α	β	δ	γ	ε	λ
	1 to 2	1 to 11	1 to r1	r1 to 2	r1 to 11	r1 to r1
ELI data	.500	.180	.320	.381	.140	.479
STEM data	.368	.116	.516	.655	.171	.174

When comparing this table with my data, we see that ELI first time freshman progress at a significantly higher rate, and repeat less. However, it is the opposite case when comparing repeating freshman. Also, ELI repeating freshman are likely to repeat their freshman year more than once, while STEM students are just as likely to leave as repeat [5]. There are numerous student populations that can be studied and modeled using Markov chains. Future work, if large amounts of data become available, could reveal interesting trends and aid in the understanding of student populations. This should lead to more informed policy-making at the University level, as well as a better-designed curriculum. With large amounts of data, this method of studying student bodies could be a very powerful tool.

The limitations of this study are largely due to the amount of data that was available. A larger sample would have given more accurate information, as well as a fuller (from freshman to graduation) picture of student paths. Part of this is due to the relatively small size of the ELI program, but also due to the fact that some information is not collected or recorded, especially in regards to students who transition from ELI to the University of Akron. It is suggested that further work be done before making decisions based on the findings of this study.

Bibliography

[1] D. P. Niranji, A. Pathirage, J. A. Morrow, D. L. Walpitage, and G. J. Skolits. Helpfulness of ESL Courses for International Students Studying in the United States. *International Education*, 43(2): 25-38, 2014.

[2] A. Phakiti, D. Hirsh, and L. Woodrow. It's not only English: Effects of other individual factors on English language learning and academic learning of ESL international students in Australia. *Journal of Research in International Education*, 12(3): 239-258, 2013.

[3] B. Song. Content-based ESL instruction: Long-term effects and outcomes. *English for Specific Purposes*, 25(4): 420-437. 2006.

[4] J.J. Higgins and S. Keller-McNulty. Concepts in Probability and Stochastic Modeling. Wadsworth Publishing Company, 1995.

[5] S. Ferko. Using A Markov Model To Analyze Retention and Graduation Rates. 2014.