Minimum-Stress Four-Year Course Scheduling via Mixed-Integer Programming

Caam378 Miniproject 3

George Lyu

November 23, 2021

Contents

1 Introduction

We create a mixed-integer program to create four-year academic schedules. The first Miniproject establishes the background and motivation:

> Rice University undergraduate students are stressed. A parade of classtime, homework, essays, exams, extracurriculars, employment, and research can leave one deprived of sleep and social health. As described by student Sophie Newman, "We [toil] away in Fondren [library], or drink 14 cups of coffee, if it means setting ourselves up for success." [1]
>
> A student's academic schedule critically influences their overall stress for the semester. Most of a student's time is spent in class or doing homework, and "students self-report academic stress as the most common reason for going to the Student Wellbeing Office or Rice Counseling Center." Though rare, students who do not take enough courses may fall off track from their academic goals. More commonly, students who take too many courses may find difficulty in deeply mastering the curriculum, accumulating "shallow knowledge in a large number of subjects [instead of] deep knowledge in a smaller number of subjects." [2] Stress created by overloaded schedules persist even after busy courses are dropped because "students may have already fallen behind in the courses that they do keep." [3] Therefore, a well-planned four-year course schedule is essential for academic success and mental well-being.

2 Model formulation

We can formulate the four-year academic scheduling problem using a mixed integer program (MIP).

2.1 Parameters

Assume we have the following known parameters:

$E =$ Existing set of courses whose credits are already obtained (e.g. via transfer credits)
$R =$ Required set of courses to complete that are not already satisfied by existing credits
$C =$ Set of required courses and their prerequisites that are not satisfied by existing credits
$F =$ Set of courses offered in fall
$S =$ Set of courses offered in spring
$D = \{1,...,8\}$ the eight academic semesters
$SA_i =$ Set of strict-and prerequisites of course i not satisfied by existing credits
$SO_i =$ Set of strict-or prerequisites of course i not satisfied by existing credits
$CA_i =$ Set of concurrent-and prerequisites of course i not satisfied by existing credits
$CO_i =$ Set of concurrent-or prerequisites of course i not satisfied by existing credits
$s_i =$ Nonnegative real number proportional to the academic/emotional stress exerted by course i

2.2 Decision variables

The decision variables represent whether each course is taken in each of the semesters.

$$\forall i \in C, \forall j \in D \qquad x_{ij} \triangleq 1 \text{ if course } i \text{ is taken in semester } j\text{; else } 0$$

2.3 Objective

The objective is to minimize the stress of the highest-stress semester.

$$\min \qquad \max_{j \in D} \sum_{i \in C} s_i x_{ij}$$

We linearize this objective using an auxiliary decision variable:

$$f \triangleq \max_{j \in D} \sum_{i \in C} s_i x_{ij}$$

Then, the objective is equal to:

$$\min \qquad f$$

We implement the decision variable f using additional constraints:

$$\forall j \in D \qquad f \geq \sum_{i \in C} s_i x_{ij}$$

In summary, we linearize our nonlinear objective as such:

$$\min \qquad f$$
$$\text{st}$$

$$\forall j \in D \qquad f \geq \sum_{i \in C} s_i x_{ij}$$

$$f \in \mathbb{R}^{0+}$$

2.4 Constraints

2.4.1 Completion constraints

Each course can be completed at most once. Required courses must be completed exactly once.

$$\forall i \in C - R \qquad \sum_{j \in D} x_{ij} \leq 1$$

$$\forall i \in R \qquad \sum_{j \in D} x_{ij} = 1$$

2.4.2 Seasonal constraints

Some classes have seasonal restrictions on enrollment. Fall-only courses and spring-only courses can only be registered in their respective seasons.

$$\forall i \in R - F \qquad x_{i1} = x_{i3} = x_{i5} = x_{i7} = 0$$

$$\forall i \in R - S \qquad x_{i2} = x_{i4} = x_{i6} = x_{i8} = 0$$

2.4.3 Prerequisite constraints

The most complicated constraints are those that enforce prerequisite requirements. Courses have strict/concurrent prerequisites and and/or prerequisites. A course cannot be enrolled in a given semester unless:

1. Each strict-and prerequisite course was completed in a previous semester.
2. At least one strict-or prerequisite course was completed in a previous semester.
3. Each concurrent-and prerequisite course was completed in a previous semester or will be completed this semester.
4. At least one concurrent-or prerequisite course was completed in a previous semester or will be completed this semester.

$$\forall i \in C, \forall j \in D \quad \sum_{\alpha \in SA_i, \beta \in D \text{ st } \beta < j} x_{\alpha,\beta} \geq |SA_i| x_{ij}$$

$$\forall i \in C, \forall j \in D \quad \sum_{\alpha \in SO_i, \beta \in D \text{ st } \beta < j} x_{\alpha,\beta} \geq x_{ij}$$

$$\forall i \in C, \forall j \in D \quad \sum_{\alpha \in CA_i, \beta \in D \text{ st } \beta \leq j} x_{\alpha,\beta} \geq |CA_i| x_{ij}$$

$$\forall i \in C, \forall j \in D \quad \sum_{\alpha \in CO_i, \beta \in D \text{ st } \beta \leq j} x_{\alpha,\beta} \geq x_{ij}$$

Some courses have corequisites: these courses must be taken in the same semester as their corequisites. Corequisite courses may be implemented as if the courses are each concurrent-and prerequisites of each other.

2.5 Summary

We summarize our MIP below:

$$\begin{aligned}
&\min && f \\
&\text{st} && \\
&\forall j \in D && f \geq \sum_{i \in C} s_i x_{ij} \\
&\forall i \in C - R && \sum_{j \in D} x_{ij} \leq 1 \\
&\forall i \in R && \sum_{j \in D} x_{ij} = 1 \\
&\forall i \in R - F && x_{i1} = x_{i3} = x_{i5} = x_{i7} = 0 \\
&\forall i \in R - S && x_{i2} = x_{i4} = x_{i6} = x_{i8} = 0 \\
&\forall i \in C, \forall j \in D && \sum_{\alpha \in SA_i, \beta \in D \text{ st } \beta < j} x_{\alpha,\beta} \geq |SA_i| x_{ij} \\
&\forall i \in C, \forall j \in D && \sum_{\alpha \in SO_i, \beta \in D \text{ st } \beta < j} x_{\alpha,\beta} \geq x_{ij} \\
&\forall i \in C, \forall j \in D && \sum_{\alpha \in CA_i, \beta \in D \text{ st } \beta \leq j} x_{\alpha,\beta} \geq |CA_i| x_{ij} \\
&\forall i \in C, \forall j \in D && \sum_{\alpha \in CO_i, \beta \in D \text{ st } \beta \leq j} x_{\alpha,\beta} \geq x_{ij} \\
&\forall i \in C, \forall j \in D && x_{ij} \in \{0, 1\} \\
& && f \in \mathbb{R}^{0+}
\end{aligned}$$

3 Collection of course data

Application of this model to create a real four-year plan requires data to provide concrete values for assumed parameters.

Data was collected from courses offered during the 2020 fall and 2021 spring semesters at Rice University [4][5]. Data was only collected for full-term courses with course level of 499 or lower (500+ level courses are typically graduate-level courses). If a course is offered multiple times (e.g. there are multiple course sessions), data was collected only for the most recent session. In total, 1,661 courses were evaluated. For each of these courses, the following data was collected:

1. Fall availability
2. Spring availability
3. Prerequisite courses
4. Corequisite courses
5. Course evaluation data (workload and grades metrics) [6]

Fall and spring availability are boolean values describing whether the specific course is offered in the respective season.

Prerequisite and corequisite courses are collected as strings. Parsing these strings into quantifiable constraints is explained further in subsection 3.1.

Course evaluation data is collected as positive numerical values from the Esther Course Evaluation page [6]. After completing a course, students are required to rate the course on a 1-5 scale for metrics such as organization, overall quality, and assignment relevance. The 1-5 average score for the workload and grades metrics were chosen to best approximate the stress of the given course. A course's stress value is equal to the product of the workload and grades metrics. Data was only collected for the 1,000 highest-enrollment courses to reduce computation load. Courses with unavailable or uncollected course evaluation data were assigned a default stress value of 3.625, which is the average stress value among collected evaluation data.

3.1 Parsing prerequisites

Of the 1,661 total courses, 548 courses have prerequisites, and 29 courses have corequisites.

Each course's prerequisite data is stored as a string of course titles and "and/or" operators. All such prerequistes are strict. Table 1 contains some examples of prerequisite strings.

Table 1: Example prerequisite strings

Title	Prerequisite string	Prerequisite type
CAAM 421	CAAM 378 AND CAAM 382	Strict-and
DSCI 301	MATH 102 OR MATH 106 OR MATH 112	Strict-or
ANTH 305	(ANTH 200 OR LING 200) AND (ANTH 301 OR LING 301)	Nested
CAAM 453	(CAAM 334 OR CAAM 335) AND CAAM 336	Nested

Our formulated model requires prerequisites written as and- or or-type prerequisites, such as that of CAAM 421 (strict-and prerequisite) or DSCI 301 (strict-or prerequisite). Our model is not useful for quantifying nested prerequisites like those of ANTH 305 and CAAM 453. We must restructure the nested prerequisites into and- or or-type prerequisites using auxiliary courses.

Algorithm 1: Parse Nested Prerequisite

Input:
prerequisite string *str* to course *crs*

while *str has a parenthetical expression* **do**
    inner ← innermost parenthetical expression ;
    Replace the entire *inner* expression within *str* with auxiliary course *auxl* ;
    Parse *inner* as a concurrent prerequiste to *auxl* ;
Parse *str* as a strict prerequisite to course *crs* ;

As example, we restructure the ANTH 305 prerequisite string: "(ANTH 200 OR LING 200) AND (ANTH 301 OR LING 301)".

1. Declare pererequisite string of ANTH 305 as: str ← "(ANTH 200 OR LING 200) AND (ANTH 301 OR LING 301)"

2. Identify "(ANTH 200 OR LING 200)" as an innermost parenthetical expression of str

3. Evaluate "ANTH 200 OR LING 200" as a concurrent-or prerequisite of new auxiliary course AUXL 100

4. Replace innermost expression in str: str ← "AUXL 100 AND (ANTH 301 OR LING 301)"

5. Identify "(ANTH 301 OR LING 301)" as an innermost parenthetical expression or str

6. Evaluate "ANTH 301 OR LING 301" as a concurrent-or prerequisite of new auxiliary course AUXL 101

7. Replace innermost expression in str: str ← "AUXL 100 AND AUXL 101"

8. str has no more innermost expressions. Evaluate “AUXL 100 AND AUXL 101” as strict-and prerequisites for ANTH 305

In summary, parsing the ANTH 305 prerequisite string decomposes the nested prerequisite into three nonnested prerequisites, allowing our formulated model to quantify the prerequisites as constraints:

1. ANTH 305 has strict-and prerequisites AUXL 100, AUXL 101
2. AUXL 100 has concurrent-or prerequisites ANTH 200, LING 200
3. AUXL 101 has concurrent-or prerequisites ANTH 301, LING 301

Parsing the 548 prerequisite strings creates 274 auxiliary courses.

Parsing corequisites is much easier. All the corequisite strings are singleton courses. We treat the course as a concurrent-and prerequisite of its corequisite, and we treat the corequisite as a concurrent-and prerequisite of the course.

3.2 Isolate courses relevant to the required course list

Parsing the prerequisite strings creates 274 auxiliary courses. In addition to the 1,661 real courses, there are now a total of 1,935 courses. Each course we evaluate in the MIP model creates about 10 constraints because one constraint is created for each semester, and additional constraints are created if prerequisites exist. Therefore, evaluating all 1,935 courses in the MIP would create a very high number of constraints, leading to cumbersome computation times.

Fortunately, very few of the 1,935 courses are relevant for a given schedule. (An applied mathematics student is unlikely to require medieval history courses.) We can reduce computation time by only evaluating the courses that are directly relevant to a student's required courses.

We designate a course as “relevant” if the course has not already been completed by the student and if at least one of the conditions are met:

1. The course is in the student's required course list.
2. The course is a prerequisite of a course in the required course list.
3. The course is a prerequisite of a course in the relevant course list.

Given a student's required course list (courses they need to complete) and existing course list (courses they have already completed), we use a modified breadth-first-search algorithm to isolate the relevant courses.

Algorithm 2: Isolate Relevant Courses

Input:
R, set of courses required to be completed
E, set of courses already completed
Output:
A, set of relevant courses

$A \leftarrow R - E$;
$Q \leftarrow$ empty queue ;
Enqueue all elements in A to Q ;
while *Q not empty* **do**
    $crs \leftarrow Q.dequeue()$;
    for *each prerequisite $prereq$ of crs* **do**
        if $prereq \notin A \wedge prereq \notin E$ **then**
            Add $prereq$ to A ;
            Enqueue $prereq$ to Q ;
return A ;

If we evaluate only isolated relevant courses, we can dramatically reduce computational load.

4 Application of model

4.1 Build a four-year plan for a B.A. in Computer Science

There are many pathways toward obtaining a Bachelor of Arts in Computer Science [7]. We select one possible pathway and apply the MIP model to create a four-year plan. We assume the student has no existing course credits. We limit the schedule to only degree requirements, so we do not account for general graduation requirements like the 120-credit-hour requirement or distribution requirements.

Inputting 17 required courses and 0 existing courses creates 25 relevant courses, 201 decision variables, and 153 constraints.

Table 2: Model-produced four-year plan for a Bachelor of Arts in Computer Science.
* annotates a course that was not in the input required course list but was added by the model to satisfy prerequisites

Semester 1 (Fall)			Semester 2 (Spring)		
Title	Hours	Stress	Title	Hours	Stress
Totals:	11	15.6	Totals:	7	15.1
COMP 140	4	6.5	COMP 182	4	10.3
MATH 102	3	6.9	MATH 211	3	4.8
ELEC 220	4	2.2			

Semester 3 (Fall)			Semester 4 (Spring)		
Title	Hours	Stress	Title	Hours	Stress
Totals:	10	14.0	Totals:	7	12.8
MATH 355	3	5.2	MATH 222*	3	3.0
STAT 310	3	3.5	COMP 321	4	9.9
COMP 215	4	5.3			

Semester 5 (Fall)			Semester 6 (Spring)		
Title	Hours	Stress	Title	Hours	Stress
Totals:	7	11.5	Totals:	8	15.8
COMP 310	4	7.1	COMP 340	4	7.6
COMP 330	3	4.4	COMP 411	4	8.2

Semester 7 (Fall)			Semester 8 (Spring)		
Title	Hours	Stress	Title	Hours	Stress
Totals:	7	14.3	Totals:	8	13.0
MATH 101	3	6.8	COMP 322	4	4.0
COMP 382	4	7.5	COMP 421	4	9.0

The produced schedule appears to satisfy our objective of minimizing stress. Stress is approximately uniformly distributed across the semesters, and no semester is substantially more or substantially less stressful than the others.

Furthermore, the COMP-department courses closely follow the recommended sequence.

Unfortunately, the MATH-department courses are unreasonably sequenced. The recommended sequence is MATH 101, MATH 102, MATH 211, MATH 221, MATH 222 (though this ordering is not enforced by official prerequisites). This ordering is not preserved in the produced schedule, and MATH 221 is not present at all.

MATH 222 was added by the model as a prerequisite, which is also unreasonable because there are easier courses that can satisfy these prerequisites. MATH 222 was likely selected as the prerequisite because it is recommended (but not enforced) to be taken alongside MATH 221, so MATH 222 individually has low stress metrics.

4.2 Build a five-semester plan to extend my own schedule to pursue a B.A. in Operations Research and a B.A. in Statistics

I have completed three semesters at Rice, and I plan to enroll in five more semesters. I hope to earn a Bachelor of Arts in each of Operations Research and Statistics [8][9]. We use the MIP model to extend my existing three semesters into a full eight-semester schedule that allows me to achieve these degrees. We also include courses that I wish to take for fun or for general graduation requirements.

Inputting 32 required courses and 39 existing courses creates 45 relevant courses, 226 decision variables, and 180 constraints.

Table 3: Model-produced five-semester plan for a Bachelor of Arts in Operations Research and Statistics, with additional courses included for general graduation requirements. Semesters 1-3 are the existing coursework; semesters 4-8 show the model-produced schedule.
* annotates a course that was not in the input required course list but was added by the model to satisfy prerequisites

Semester 1 (Fall)			Semester 2 (Spring)		
Title	Hours	Stress	Title	Hours	Stress
Totals:	16	22.5	Totals:	14	24.3
ARTS 165	3	2.2	COLL 145	1	1
COMP 140	4	6.5	COMP 182	4	10.3
ECON 100	3	3.7	ECON 307	3	3.0
MATH 221	3	5.3	FWIS 188	3	7.0
MECH 202	3	4.8	MATH 222	3	3.0

Semester 3 (Fall)			Semester 4 (Spring)		
Title	Hours	Stress	Title	Hours	Stress
Totals:	13	18.7	Totals:	9	17.3
BUSI 296	3	2.9	MATH 112*	3	7.1
CAAM 378	3	6.0	CAAM 382	3	4.2
COMP 215	4	5.3	CLAS 235	3	6.0
MATH 354	3	4.5			

Semester 5 (Fall)			Semester 6 (Spring)		
Title	Hours	Stress	Title	Hours	Stress
Totals:	13	16.5	Totals:	16	16.7
DSCI 301*	4	3.6	DSCI 302*	3	5.0
CAAM 471	3	3.6	MUSI 117	3	1.9
STAT 418	3	5.6	STAT 410	4	4.3
STAT 482	3	3.6	STAT 405	3	2.9
			MATH 302	3	2.5

Semester 7 (Fall)			Semester 8 (Spring)		
Title	Hours	Stress	Title	Hours	Stress
Totals:	10	16.2	Totals:	12	16.5
COMP 449	4	6.4	DSCI 304	3	2.7
STAT 413	3	4.2	DSCI 305	3	3.0
CAAM 335*	3	5.7	STAT 419	3	6.5
			STAT 421	3	4.3

Again, the program effectively smooths the stress across the five semesters. The produced schedule demonstrates the feasibility of an OR/STAT double-major, which gives me confidence that this academic pursuit is achievable. Furthermore, the sequence of CAAM and STAT courses appears reasonable. The addition of DSCI 301 and DSCI 302 as prerequisites to necessary courses also makes sense because these two courses are necessary and optimal prerequisites to DSCI 304.

Unfortunately, the model again assigns extraneous prerequisites. Specifically, the model adds courses MATH 112 and CAAM 335 as prerequisites to required courses. These additional courses do not make sense–most professors will waive prerequisites of MATH 112 and CAAM 335 if the student has taken MATH 222 and MATH 354. However, this precedent of prerequisite-waiving is not not officially expressed in the prerequisite data, so it does not manifest in the model-produced schedule.

4.3 Improvement opportunities

Many course sequencing issues could be improved by also accounting for recommended prerequisites. In addition to official prerequisites, some courses have recommend prerequisites, comprising of courses that the professor(s) deem useful but not completely necessary for completing the course. Most STEM courses sequence based on recommended prerequisites instead official prerequisites, so recommended prerequisites may provide for more accurate sequencing.

However, accounting for recommended prerequisites will introduce challenges. Parsing recommended prerequisites is more challenging than parsing official prerequisites because recommended prerequisites are written into the body of the course description, whereas official prerequisites are written in their own dedicated field. Additionally, recommended prerequisites have varying utility: sometimes recommended prerequisites contain useful and necessary coursework, but other times recommended prerequisites can be ignored without consequence. Distinguishing between useful and overly restrictive recommended prerequisites will be challenging.

Another area the model can be improved is by reducing single-subject load. Students are discouraged from taking many courses in the same subject area in a single semester, as single-subject load can quickly lead to burnout.

A second objective could be added that maximizes course enjoyment. Though stress-minimization is important, a particularly enjoyable/impactful class may provide enough fun to outweigh the difficulty.

Finally, the model should account for heterogeneity in semesters. Some semesters are particularly stressful: the first semester in freshman year is very stressful because freshman are transitioning from high school to college. It is also recommended that the last two semesters in senior year should be scheduled lightly because seniors cannot easily drop their courses, as there are very few opportunities to complete any dropped courses. Semester heterogeneity can be implemented by creating a semester-based stress multiplier, where the first semester's and senior semesters' stresses can be multiplied by some constant.

Unfortunately, many common scheduling considerations, like professor-endorsed prerequisite overrides, time-based scheduling conflicts, and class size restrictions, cannot be easily quantified or implemented.

References

[1] Sophie Newman. “Stress exists in the future, so be present”. In: *The Rice Thresher* (Mar. 2015). URL: https://www.ricethresher.org/article/2015/03/stress-exists-in-the-future-so-be-present.

[2] Emily Abdow. “Faculty Senate approval of credit hour limit in face of student protests spurs campus debate”. In: *The Rice Thresher* (Apr. 2016). URL: https://www.ricethresher.org/article/2016/04/faculty-senate-cuc-vote.

[3] David Leebron and Marie Lynn Miranda. “President Leebron, Dean Miranda address credit hour cap protests in email to student body”. In: *The Rice Thresher* (Apr. 2016). URL: https://www.ricethresher.org/article/2016/04/leebron-miranda-student-email.

[4] Rice University. *Course Schedule Fall Semester 2020*. 2021. URL: https://courses.rice.edu/courses/!SWKSCAT.cat?p_term=202110.

[5] Rice University. *Course Schedule Spring Semester 2021*. 2021. URL: https://courses.rice.edu/courses/!SWKSCAT.cat?p_term=202120.

[6] Rice University. *Esther Course and Instructor Evaluation Display*. 2021. URL: https://esther.rice.edu/selfserve/swkscmt.main.

[7] Rice University. *Bachelor of Arts (BA) Degree with a Major in Computer Science*. 2021. URL: https://rice-preview.courseleaf.com/programs-study/departments-programs/engineering/computer-science/computer-science-ba/#requirementstext.

[8] Rice University. *Bachelor of Arts (BA) Degree with a Major in Operations Research*. 2021. URL: https://rice-preview.courseleaf.com/programs-study/departments-programs/engineering/operations-research/operations-research-ba/#requirementstext.

[9] Rice University. *Bachelor of Arts (BA) Degree with a Major in Statistics*. 2021. URL: https://rice-preview.courseleaf.com/programs-study/departments-programs/engineering/statistics/statistics-ba/#requirementstext.