Course Difficulty Calculation

Andrew Dumit, Qiwei Li, Haomiao Meng

Abstract

In this paper, we propose a framework to calculate course difficulty index (CDI), which is based on the difference between student's performance in the course and student's overall academic strength. We use three specific approaches to illustrate this framework, including two definitions of course difficulty (GPA penalty and ranking penalty) and one algorithm (least squares method). Finally, we apply these three approaches to the University of California, Davis (UC Davis) Undergraduate Database to compute CDI for each course and visualize the results for course comparison.

1 Introduction

It is widely understood that there is a large disparity between the difficulty of different courses. Nevertheless, it is not understood how different courses are difficult. Some courses are difficult because they take a lot of time, while others are difficult because of the teacher. Though Universities do not understand course difficulty, students attempt to take advantage of it regularly. However, course difficulty is important for universities to understand if they wish to maximize student performance. At UC Davis, one method to identify course difficulty is for students to report their perception of course difficulty at the end of each quarter. However, it has been shown repeatedly that this perception is heavily biased. There is evidence that course ratings are biased by professor attractiveness, ethnicity, and gender (Riniolo *et al.* 2015, Macnell *et al.* 2014, Anderson *et al.* 2007). Further evidence shows that these ratings are also biased by grading standards, expected grade, and the halo effect (Keeley *et al.* 2013, Zabaleta 2007, Ducette *et al.* 1982, Addison *et al.* 2006).

Another way to access course difficulty is to use grade of each student in that course. This is a more objective method since grades, in most cases, should reflect how well a student did in the class and students should tend to have a better grade in easy courses. However, considering different students have different levels of incoming ability or strength, it is reasonable to try and control for the individual effect of each student. So, another way to acquire course difficulty is by comparing the performance of a student in class with strength of that student.

In this paper, we propose several methods to measure student's class performance and overall strength and ultimately calculate course difficulty. The course difficulty measures we propose should be useful to advisors looking to help students select courses that will maximize those students' performance as well as for universities that wish to provide equal opportunity for students. Furthermore, we hope this measure of course difficulty will help in predicting student success rate in combination with other indicators.

2 Methodology

The overall goal of this paper is to introduce a framework to compute course difficulty index (CDI). That framework is based on the following formula:

$$\text{Course Difficulty Index} = \text{Student's Class Performance} - \text{Student's Overall Strength} \quad (1)$$

Equation (1) is the general model for course difficulty. By looking at its right hand side, it actually reflects one student's opinion about one specific course. If we aggregate all student's opinion about that course, then we can get this course's difficulty index. Note that we do not define "studnet's class performance" yet, it can be any reasonable measure. A student's overall strength can be seen as the student's innate ability, which must be comparable to student's class performance. For example, if student performance is on a scale from 0 to 4, i.e. GPA, student strength must also be on such a scale. However, since grade point is the only objective, accessible source, all the approaches will be based on grade point in this paper. The following sections give three typical methods within this framework, including two definitions of course difficulty and one algorithm to evaluate it.

2.1 GPA Penalty

One definition for course difficulty is GPA penalty, meaning that hard courses will impose more penalty on student's GPA. Following University of Michigan's study on GPA penalty (Koester 2015), we can define student's class performance and student's overall strength in equation (1) as

$$\begin{aligned}\text{Student's Class Performance} &= \text{grade point in this course,}\\ \text{Student's Overall Strength} &= \text{grade point average in other courses,}\end{aligned}$$

then the course difficulty index (CDI) could be written as

$$CDI = GP - GPAO.$$

This definition is based on a simple idea. If a student feels one course is harder than others, he or she tends to have a lower grade, compared to other courses, thus a negative opinion will be contributed to the overall course difficutly measurement. It is called GPA penalty because difficulty of course is entirely based on student's absolute grades, regardless of each student's relative performance to his or her classmates. If students get better grades in one course, then that course is thought to be an easy one.

After computing the difference for each student, the difficulty index D_j for course j is

$$D_j = \frac{\sum_{i=1}^{N_j}(GP_{ij} - GPAO_{ij})}{N_j}, \tag{2}$$

where N_j is the number of students taking course j, GP_{ij} and $GPAO_{ij}$ are grades student i got in course j. Notice that here GPAO could be also replaced by GPA since basically both are measurements regarding student's strength.

2.2 Ranking Penalty

Consider the case where a student earned an A- in a course, which is objectively close to the best possible grade, yet every person in that course earned an A. If the student's GPAO is a B, then a positive feedback (that course is easy) will be added to the course difficulty index according to the first definition. However, a closer examination will reveal that the student actually did worse than everyone else in this class. If we believe every student will work hard in his or her class, then student's class rank may better reflect the true difficulty of that course.

This brings us to the second definition of course difficulty, ranking penalty. Ranking penalty follows the assumption that hard courses will impose more penalty on student's class ranking. For ranking penalty, instead of using the real grade point, we use the normalized grade point to evaluate student's class performance. Recall equation (1), we have

$$\text{Student's Class Performance} = \text{normalized grade point in this course,}$$
$$\text{Student's Overall Strength} = \text{normalized grade point average in all courses.}$$

After computing the difference, we actually have the newly defined course difficulty. In order to illustrate the extension of this framework, here we impose sign function (1, -1 or 0) to the difference to define course difficulty index, namely,

$$CDI = \text{sign}(NORM\_GP - NORM\_GPA).$$

$NORM\_GP$ can be seen as percentile of student's grade in this course and $NORM\_GPA$ is simply averaging $NORM\_GP$. More specifically, let NGP_{ij} denote the normalized grade point of student i in course j and $NGPA_i$ be the overall normalized GPA of student i, then

$$NGP_{ij} = \frac{GP_{ij} - \overline{GP}_{.j}}{\text{std}(GP_{.j})}, \quad NGPA_i = \overline{GP}_{i.}.$$

By aggregating all student's performance, the difficulty index D_j for course j is

$$D_j = \frac{\sum_{i=1}^{N_j}\text{sign}(NGP_{ij} - NGPA_i)}{N_j}, \tag{3}$$

where N_j is the number of students taking course j.

The sign(.) function here is to let student vote for difficulty level of a course. If a student receives a worse class rank, compared to his overall average ranking, he or she would give a down vote. Then one benefit of the sign function is that it can control the range of course difficulty since by averaging voting, the index must be a number between -1 to 1. Note that the sign function can be replaced by any other one-to-one mappings.

2.3 From ANOVA Point of View

The previous two sections offer two definitions of course difficulty. In this section, we will focus on the algorithm to compute course difficulty. For the purpose of illustration, we still use the idea of GPA penalty to define course difficulty, namely, the computation is based on student's real grade point.

A paper from Carnegie Mellon University (Caulkins *et al.* 1996) has used an additive model to compute student's ajusted GPA and course difficulty, which was, in principle, an ANOVA model for student's grade point. It breaks down GP_{ij}, grade of student i in course j, into two parts, student's attribute $AGPA_i$ and course's attribute D_j,

$$GP_{ij} = AGPA_i + D_j + \epsilon_{ij}, \tag{4}$$

where ϵ_{ij} is the error term, which is assumed with mean 0 and uncorrelated with $AGPA_i$ and D_j.

Model (4) can be interpreted as this way: each student's course grade can be explained by two factors, one is student's own "innate ability" and the other is the effect of the course on real grade point. Then $AGPA_i$ can be seen as adjusted GPA of student i and D_j is difficulty of course j. Solve this problem by least squares method (LS) and we arrive at:

$$\min \sum_i \sum_j (GP_{ij} - AGPA_i - D_j)^2.$$

The first order condition (FOC) gives us

$$AGPA_i = \frac{\sum_{j=1}^{n_i}(GP_{ij} - D_j)}{n_i}, \quad D_j = \frac{\sum_{i=1}^{N_j}(GP_{ij} - AGPA_i)}{N_j}, \tag{5}$$

where n_i is the number of courses student i has taken and N_j is the number of students taking course j. An iteration procedure would solve this set of equations.

Comparing equation (2), (3) and (5), we immediately notice that all D_js have a similar form. In fact, if we rewrite equation (4) as

$$D_j = GP_{ij} - AGPA_i - \epsilon_{ij},$$

then we can define

$$\text{Student's Class Performance} = \text{grade point in this course},$$
$$\text{Student's Overall Strength} = \text{student's adjusted GPA},$$

thus the course difficulty index would be

$$CDI = GP - AJUST\_GPA,$$

which is exactly the same form as the previous methods. Furthermore, if we use the idea of ranking penalty, we could use $NORM\_GP$ to replace GP. Then $AJUST\_GPA$ would be $AJUST\_NORM\_GPA$. We can also impose a one-to-one function to transform the inputs such that the result will be bounded.

3 Data

UC Davis Undergraduate Database provides information for close to 130,000 students enrolled from 2000 to 2015. Among other factors, the data contained in the database holds records for the grades of each student in over 6,000 different courses ultimately adding up to over 4 million observations. Since the difficulty of one course may vary over years, even taught by the same instructor, we treat the unique combination of course, instructor and term as a unique course. This leads to a total of 100,000 unique courses over 15 years.

We use the traditional letter grade conversion, namely, A=4, B=3, C=2, D=1 and F=I=NG=Y=0. The grade may also be modified by a plus (+) or a minus (-), which adds or subtracts .3. Since it is unclear how to compare PASS/FAIL grades with letter grades due to the fact that a PASS $\neq$ A,B or C, we discard student records with PASS/FAIL courses among all registered students. We also exclude records with NA grades, which is typically due to withdrawal, in progress, or for other unknown reasons. Furthermore, for those classes with only one student, we define $NORM\_GP$ to be 0 to avoid invalid standard error.

4 Results Visualization

There are several ways to visualize course difficulty index. First, since the same course across different terms is regarded as different courses, we could look at time series of one course. At UC Davis, it is often assumed that summer courses are easier than their regular term counterparts. Figure 1 shows the difficulty changes of Nutrition 10 (NUT 10). It is very clear that the summer courses are much easier than normal term courses, confirming the widely held belief.

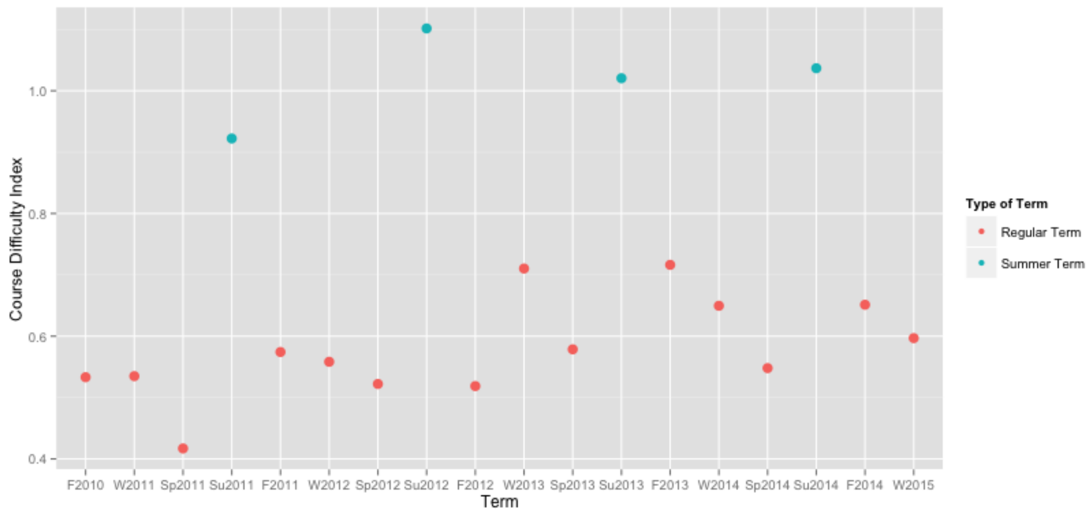

Figure 1: mean course difficulties of Nutrition 10 from 2011 to 2015.

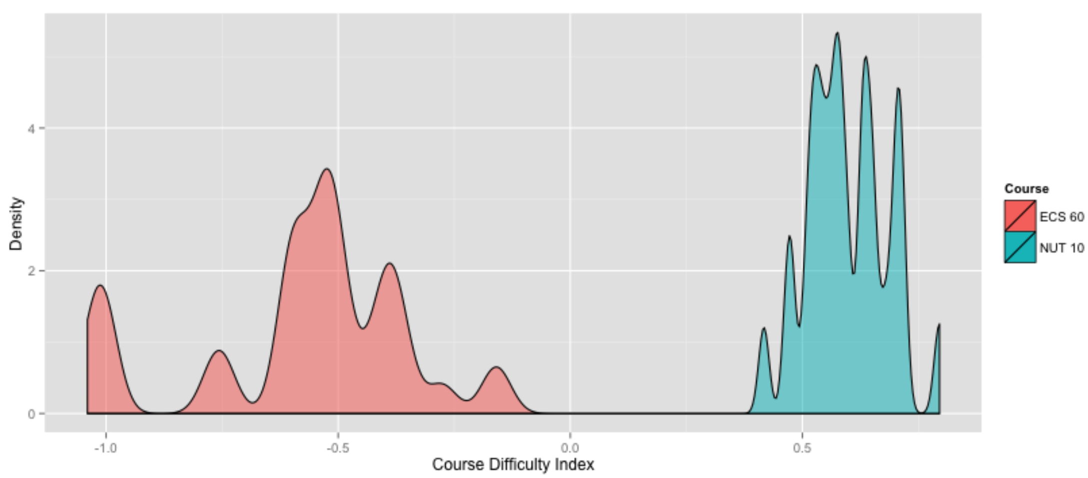

Figure 2: density of course difficulties of Nutrition 10 vesus Computer Science 60.

By assembling course difficulty index across terms and instructors, we can plot the density of these indices and compare among different courses. Among students, NUT 10 is rumored to be an "easy A" and a "GPA booster". Conversely, Computer Science 60 (ECS 60), is believed to be one of the more challenging courses. Figure 2 is an example of showing the overall difficulty of NUT 10 and ECS 60 with summer terms removed for consistency. From figure 2, ECS 60 is more difficult than NUT 10, which meets our expectation.

By equation (1), we can compare student's class performance and student's overall strength individually and construct a plot to express it. If each student's class performance is exactly equal to his or her strength, the plot will be a 45 degree line. By comparing how students' performances deviate from the 45 degree line, we can see how much the GPA penalty or ranking penalty for the course affects the students in that course and how a particular trait affects the grade penalty. Figure 3 shows the GPA penalty for Computer Science 30 (ECS 30) with a gender split. This split appears to show that males tend to do better in this class than females.

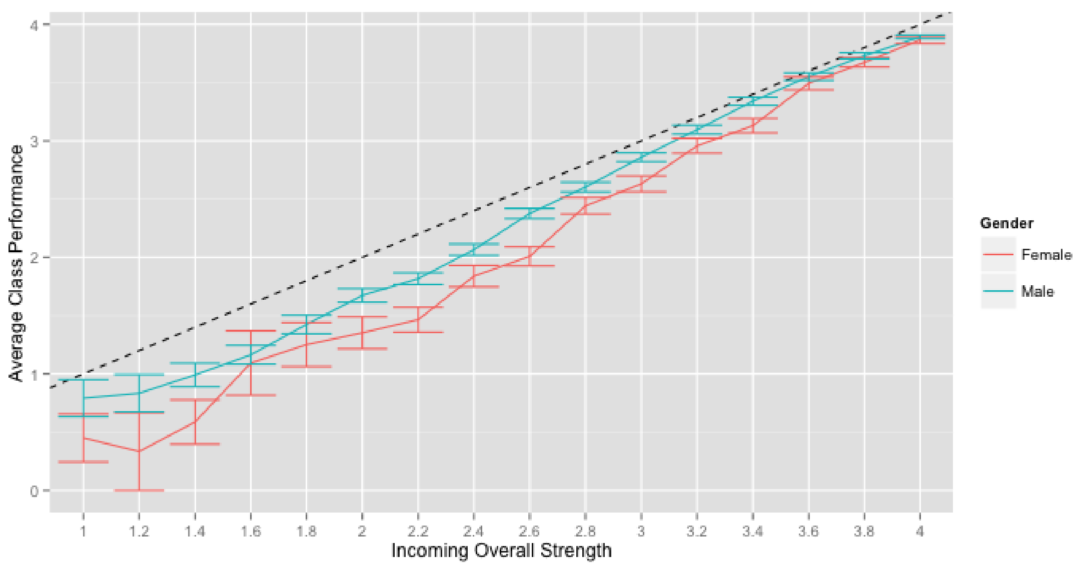

Figure 3: Grade penalty of Computer Science 30 split by gender.

According to equation (2), (3) and (5), the course difficulty can be calculated by aggregating all the student's opinion. But if we sum up these differences by gender, then we could actually have male prefered courses and female perfered courses. Similar to how Figure 3 showed that ECS 30 is preferred by males, Figure 4 shows the 32 highest enrolled courses' gender preferences at UC Davis. Similarly, we could also aggregate by ethnicities or first generation status for further study on course preference.

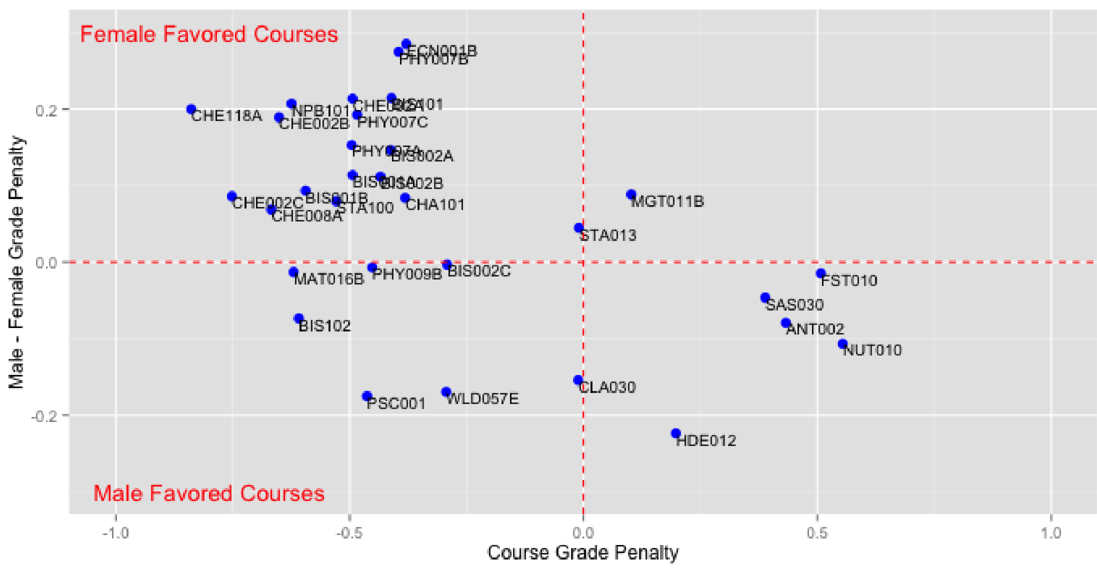

Figure 4: Overall grade penalties compared to gender separated grade penalties of all 32 courses with over 500 student enrollment.

5 Conclusion

In this paper, we proposed a framework for calculating a measure of course difficulty, the course difficulty index. This framework relies on the relationship between the grade penalty of a course and the students relative academic strength. Within this framework, we implemented two different methods for calculating the course difficulty of UC Davis courses and one algorithm for solving for the course difficulty index. Although there are an uncountable number of methods that could be implemented within this framework, the ones proposed in this paper are the ones that seem to be the easiest to understand and convey to administrators. The ease with which one can convey the core meaning of the indicies is an important feature of the methods if there is to be any practical use for the indicies.

Then, using the calculated indicies, we demonstrated their usefulness in identifying the easier and more difficult courses, and which courses are preferred by certain subsets of students. With this knowledge, we hope advisors can recommend more appropriate course loads for students and action can be taken to identify the meaning behind some of the differences in grading penalty for subsets of students for which there theoretically should be no gap.

References

[1] Zabaleta, Francisco. "The Use and Misuse of Student Evaluations of Teaching." *Teaching in Higher Education* 12.1 (2007): 55-76. Web.

[2] Riniolo, TC, KC Johnson, TR Sherman, and JA Misso. "Hot or Not: Do Professors Perceived as Physically Attractive Receive Higher Student Evaluations?" *General Psychology* 133.1 (2006): 19-35. *Pubmed.gov.* NCBI. Web. 30 June 2015.

[3] Ducette, Joseph, and Jane Kenney. "Do Grading Standards Affect Student Evaluations of Teaching? Some New Evidence on an Old Question." *Journal of Educational Psychology* 74.3 (1982): 308-14. Web.

[4] Macnell, Lillian, Adam Driscoll, and Andrea N. Hunt. "Whats in a Name: Exposing Gender Bias in Student Ratings of Teaching." *Innovative Higher Education Innov High Educ* (2014): n. pag. Web.

[5] Addison, William E., John Best, and John D. Warrington. "Students' Perceptions of Course Difficulty and Their Ratings of the Instructor." *College Student Journal* 40.2 (2006): n. pag. Web.

[6] Anderson, K. J., Smith, G. (2005). Students preconceptions of professors: Benefits and barriers according to ethnicity and gender. *Hispanic Journal of Behavioral Sciences, 27,* 184-201. doi:10.1177/0739986304273707

[7] Keeley, J. W., T. English, J. Irons, and A. M. Henslee. "Investigating Halo and Ceiling Effects in Student Evaluations of Instruction." *Educational and Psychological Measurement* 73.3 (2013): 440-57. Web.

[8] Elliott, R. and Strenta, A. C. 1988 Effects of Improving the Reliability of the GPA on Prediction Generally and on Comparative Predictions for Gender and Race Particularly, *Journal of Educational Measurement*, Vol. 25, No. 4, pp. 333-347.

[9] Caulkins, Jonathan P., Patrick D. Larkey, and Jifa Wei. "Adjusting GPA to Reflect Course Difficulty." *CMU Research Showcase* (1996): n. pag. Web.

[10] Koester, Ben. "Grade-Based Performance Prediction." SLAM 2015. Michigan, Ann Arbor. *SLAM.* Web.