Math 381 Justin Harding, Philip Placek
Project February 22, 2010

A Scheduling Model with Advanced Constraints

Abstract

The UW computer science department runs a computing lab for students enrolled in introductory courses. Each quarter, approximately 45 Teaching Assistants (TAs) are scheduled for hours throughout the week in the lab. The current process for assigning hours is a seniority based first-come-first-serve sign up. This paper will use methods of linear programming to optimize this scheduling process. We then look at ways to further improve the scheduling by giving TAs better hours based on their seniority and by assigning TAs based on when homework is due.

Contents

3 Introduction

The Introductory Programming Lab (IPL) is a study center where students can get help with homework for the computer science department. The current system of assigning hours for TAs leads to the outcome where many younger TAs get stuck with bad hours, and times when less experienced TAs must staff more difficult hours. The goal of this project is to determine the most effective way to schedule hours for the IPL so that as a whole, the TAs get hours they are happy with. Also we want to make sure that the IPL is staffed according to the number of students that visit.

Our objective is to maximize the TAs' "happiness". That is, we wish to schedule TAs so that as many as possible will work the hours that they choose. TAs must work two or three hours per week, and they are allowed to split up their hours or work in one big block. The constraints we are facing are those of a standard scheduling problem: people can only work at times which are convenient for them and each hour requires a certain number of TAs [Chvátal].

The IPL lab uses a "tracking system" which keeps record of how many students use the lab and at what times. This data will be used to determine how many TAs should be staffed at given hours (times at which a lot of students use the lab need more TAs). Currently, the number of TAs is based on their own estimate of when students use the lab and not by data records. A similar case can be found in Berman *et al.*, where scheduling was based on their estimates of consumer demand [Berman et al].

It is our hope that this project will produce a working model that can be used to schedule TAs in the future, or at least be used as a basis for scheduling decisions. It could be used by updating the values of hours and processing information from the IPL tracking system. Much of this process could be automated through a data mining computer program to facilitate use. While our current model is in no way complete, it provides the foundation upon which a more advanced model can be built.

4 Data Collection

Our data collection came in the form of a survey, which was distributed by email to all 45 of the TAs. In the survey, each TA was asked to rank their hours (for each day of the week) of work by preference on a scale of 0 to 3, with 3 being the hours they preferred to work the most, 2 for hours that are good, 1 for hours they could work but prefer not to, and 0 for hours they cannot work. The survey conducted has each TA rate their preferences for each half hour of time. Since TAs must work in hour chunks, the code will need to make sure that TAs are scheduled for at least an hour at a time.

Each TA was also requested to submit how many quarters they had previously worked. This would give us an indication on which TA's preferences should be prioritized based on their seniority. The course listing on the university website lists the TAs and what class they teach (either 142 or 143). However in the future the survey could be made to include which section the TA teaches.

Twenty out of the forty five TAs responded to the survey. In order to fill in the remaining gaps, a monte carlo simulation was used to randomly assign preferences to the TAs. Since all the TAs are students, we assumed their preferences would be similar. So if thirty percent of the TAs who responded listed hour 1 with a preference of 3, then each of the random TAs would have a thirty percent chance of listing hour 1 with a preference of 3. The seniority level was also randomized for these TAs in the same manner. Both the TAs who responded to the survey and the ones that were assigned random preferences were used in conjunction in determining the optimal schedule.

Information on what days homework is due can be easily gathered from the TAs. Typically, CSE 142 homework is due on Tuesday night while 143 is due on Thursday night. We will further optimize our schedule by making sure there's always a 142 or 143 TA working their respective night when homework is due.

5 The Initial Model

The basis for our model is similar to that which Pinedo outlines as a shift schedule [Pinedo]. That is, there are a certain number of personnel required for each time period, and each employee is assigned a certain shift pattern. In our model, the shift pattern is simply working for either one or two hours at a time. Since this is relatively simple we do not need to account for different shift patterns in our model and we will not need to create an extra variable to reprsent shift patterns as Pinedo does in his model.

The solution strategy used for this problem will be to model the schedule as a linear program. Ultimately, this is an integer program as we cannot have half a TA working for $\frac{1}{5}$ of an hour. This paper will examine several different options, such as focusing solely on seniority, solely on good staffing, and a combination of the two. First we will solve the problem without adding any of the additional constraints previously mentioned.

5.1 The Objective Function

The decision variables for the model are whether a TA works a given hour (these variables will either be 0 or 1). The objective function weights each hour a TA works by the score value they indicated on the survey (1 through 3). The IPL lab is open from 10:30am to 9:30pm Monday through Saturday, giving a total of 22 half hour chunks per day, leading to a total of 132 half-hours that are rated. The objective function looks like:

$$MaxZ = \sum P_y * TA_{xy} \tag{1}$$

where P_y is TA_y's preference level, and x are the hours. TA_{xy} takes on the value of either 0 or 1. Since there are 45 TAs total, the values of y range from 1 to 45, and x ranges from 1 to 132. The linear program is a maximize problem since the goal is to give TAs as many of their preferences as possible.

5.2 Constraints

The first constraint needed is making sure that all TAs work exactly two hours. To allow for more flexibility, we allowed the TA's to work up to three hours. Each TA has a constraint looking like:

$$\sum_{x=1}^{132} TA_{x1} \leq 6 \tag{2}$$

It should be noted that each variable is only listed in the constraints if the TA rated it with a preference not equal to zero. If TA 1 ranked the first hour with a zero, then TA_1HR_1 would not be listed in the above constraint. This rule applies for the entire model.

The next constraint involves making sure there is the right number of TAs working each hour. Our model allows for an extra TA to be working at any give time, so 1 must be added to any constraint. For example if 2 TA's are needed for the eleventh hour the constraints would look like:

$$\begin{aligned} \sum_{y=1}^{45} TA_{11y} &\geq 2 \\ \sum_{y=1}^{45} TA_{11y} &\leq 3 \end{aligned} \tag{3}$$

Finally, upper bounds need to be placed on the variables. If TA 1 can work hour 1 (in other words they rated the hour with a value above zero), but cannot work hour 2 (they rated hour 2 with a value of zero), the constraints would be:

$$TA_1HR_1 \leq 1$$
$$TA_1HR_2 \leq 0$$

Of course, as with almost any linear program, all the variables must be greater than or equal to zero. These implicit constraints were included in the model.

6 Compiling

The survey conducted reports all the results into an excel spreadsheet. The first step to finding the solution is to put the data into a linear program in cplex format. To do so, a matlab program was written to take the data and put it in a text file in cplex form. The first part of program writes the objective function using the preferences indicated. Then the program writes the constraints making sure TAs work two hours and the required number of TAs are working the given hour. Finally, the upper bounds are added to the constraints, and the implicit constraints (all variables greater than or equal to zero) are added as well.

Once the text file is generated, glpsol is used to solve the linear program. Since the output file produced by glpsol is incredibly huge due to the number of variables, a java program is used to piece out the important information. The program runs through the output file and finds the times where the TAs work and produces an easy to read schedule.

7 Modifying the Model

One of the original goals of the model was to apply levels of seniority when determining which TAs get preferenced over the others. The last question in the survey asks how many quarters the TA has worked in the past. This number is simply multiplied in the objective function so that TAs with lots of experience have more weight in getting their top picks. We let E_y be the experience in number of quarters for TA_y, so equation (1) becomes:

$$MaxZ = \sum E_y * P_y * TA_{xy}$$

The second constraint planned for this model is to make sure a 142 TA is always working on Tuesday, when 142 homework is due, and likewise for Thursday with 143. In his paper, Bard has a similar problem with his model, in that he has two type of employees. In order to solve that problem, Bard created two separate variables to represent the two types of employees [Bard]. We will use his model as a guide in solving our problem.

Currently there are 24 TAs for 142, and 21 TAs for 143. The first 24 TAs in the program represent 142 TA's, and the remaining 21 represent 143 TAs. Additional constraints are added in of the form:

$$\sum_{y=1}^{24} TA_{30y} \geq 1$$

So for hour 30, there must be at least one TA working from the first 24 TAs. Since all these TAs are 142 TAs, the requirement is satisfied. This is applied to all the hours on Tuesday and Thursday.

After running our model and finding a solution, we came across a problem with the result. An optimal solution was found, but TAs were not placed in hour blocks, and were instead only placed by half hours. In order to solve this problem, additional start time variables are created in the objective function and additional constraints are added. Equation (1) now looks like:

$$MaxZ = \sum E_y * (P_y + P_{y+1}) * TA_{xyz}$$

The values in the parenthesis represent the preference values, and z represents the start time. In our model, we allowed three start times to allow for more flexibility in creating the optimal schedule. Additional constraints are needed to ensure that the start 2 value does not overlap with the first hour (in other words, the start time is not 30 minutes after the first start time). Also the two start times must not be the same. For each x, y, and z, the constraints look like:

$$TA_{xyz} + TA_{xy(z+1)} + TA_{xy(z+2)} + TA_{x(y+1)(z+1)} + TA_{x(y+1)(z+2)} \leq 1$$

8 Results & Future Implications

In solving the initial model before the updates, a feasible solution was obtained. At this point we only used the values given from the survey, and did not use any randomly generated TAs.

After updating our model with randomly generated TAs, we solved the LP and found a feasible solution.

Although we made mention of the tracking system used to keep record of how many students show up to the lab, we did not use it in our model. The current tracking system does not keep track of the actual time of each question, so it was not as directly applicable. In the future, by adding this to what the system keeps track of, we could get a better estimate of what the TA requirements are for each hour.

We used three models to figure out what objective functions worked well for the scheduling. Results (1) shows the scheduling chart where the objective values were multiplied by the unmodified experience level. Results (2) shows the schedule with no experience weighting, thus maximizing overall happiness. Finally, Results (3) is a more balanced schedule, where each decision variable is multiplied by the square root of the experience. This still favors older TAs, but will give younger TAs a better chance of getting their favorite hours.

Our current method of establishing preferences is a simple linear ranking scale (TAs rate their hours 0-3). One way to improve upon this model would be a more advanced ranking system. For instant a TA may only prefer to work a certain hour if they are working the hour before it. Further improvement on the survery would allow our model to be expanded.

Our current method of establishing preferences is a simple linear ranking scale (TAs rate their hours 0-3). One way to improve upon this model would be a more advanced ranking system. For instant a TA may only prefer to work a certain hour if they are working the hour before it. Further improvement on the survery would allow our model to be expanded.

Another problem to solve would be the “swiss cheese” schedule problem. This occurs if a TA works for an hour, has a hour break, and then comes back to work for another hour. This scenario is probably not desireable for most people, so further constraints would be needed to ensure this does not happen.

9 Relaxations for Non-Feasible Solutions

With future use, it is possible that there will be no optimal solution based on the data given. Either TA preferences could conflict, or there could be too much or too little staffing to fill the IPL fully.

In order to ensure the model works, there are several places the code could be modified. A first, more ideal solution would be to change the number of shifts people work. In such a case, the code could be modified to allow 4 start values. If there are time slots where no one can actually work, the constraint requiring certain staffing could be loosened. By including a relaxation variable in the minimum staffing constraint, we could allow for staffing deficiency. This would still staff all required hours, but not as fully as the current model does.

This relaxation would be most effective in Summer, where the number of TAs decreases to 10. There would be a higher chance of required hours with conflicts. By taking these into account, the model should function for the foreseeable future.

10 Conclusion

Overall, the resulting schedules were very similar. We can see slight differences in which experienced TAs get their higher preferred slots in Results (1) than in the other two models. Because all the results are feasible, any one of the solutions would work. Ultimately, the model implemented will depend on the TA coordinators or instructors.

The aim of this project was to provide a model that the IPL lab can use to e?ciently schedule TAs. By allowing TAs to rank their preferences on the hours they want to work, we were able to use a linear program to find the optimal schedule that the TAs would be happy with. Additionally, we modified the model so that TAs who had more experience were given got seniority their preferred hours, and so that there is always a TA working on the day homework was due for the respective classes. This model provides a basic template for a schedule, which can easily have manual modifications made if necessary.

Keeping in mind the limitation that the schedule might have few places where TAs have gaps in their hours, a few emails at the beginning of the quarter can sort out the rough patches in the model. If this is adopted for use, we could take information from TAs to establish how pleased they are and continue to refine the model.

References

[Chvátal] Chvátal, Vašek. Linear Programming. McGill University, W.H. Freeman and Company. 1983.

[Berman et al] Berman, Oded; Larson, Richard C.; Pinker, Edieal. Scheduling workforce and workflow in a high volume factory. Management Science, Volume 43. 1997.

[Bard] Bard, Jonathan F. Staff scheduling in high volume service facilities with downgrading. IIE Transactions, Volume 34. 2004.

[Pinedo] Pinedo, Michael L. Planning and Scheduling in Manufacturing and Services. Springer Series in Operations Research. 2005.