10th International Conference of the Practice and Theory of Automated Timetabling
PATAT 2014, 26-29 August 2014, York, United Kingdom

Timetabling in Higher Education: Considering the Combinations of Classes Taken by Students

Carlos A. Sánchez

Keywords: Higher education, timetabling, association rules, data mining, analytics, market basket analysis, graphs, network analysis

1 Introduction

This work is part of a project that aims to develop an analytics based architecture and methodologies to support the design and implementation of collaborative timetabling systems in higher education. As part of the larger project, this contribution will report on methodologies and algorithms that are being developed to enable the exhaustive identification of combinations of courses that students take from the offered class schedule and combinations of interest that are frequently not possible due to schedule conflicts. This article discusses the architectural components that identify the course offerings that limit the enrollment options for students based on data on schedules and enrollments from recent previous terms.

An aspect that has not been considered in detail in the higher education timetabling literature, relates to the inefficiencies embedded in the constraints that are specified and passed to optimization algorithms. A common approach to timetabling in higher education is to take the requirements and constrains as a given, and then to use optimization algorithms to search for optimal solutions that meet those requirements and constraints. The problem is that optimization algorithms are neither designed nor intended to identify and solve inefficiencies embedded in constraints passed to them.

With the previous ideas in mind, the goal of this work is to help scheduling authorities gain a better understanding of enrollment patterns, identify the schedule offerings that frequently limit the enrollment options for students, and determine the sections that need to be offered that are free of schedule conflicts. The ideas proposed in this work do not require a detailed knowledge about the programs offered at the institution and provide information on courses that are good candidates for section offerings that should be free of schedule conflicts. This information would be intended to help scheduling authorities to produce

School of Information Sciences, University of Pittsburgh
Pittsburgh, PA, 15260, USA
carlosal@pitt.edu

class schedules that better reflect the needs and interest of the students, the institution, and when applicable specify better informed constraints to be passed to timetabling algorithms.

This initiative is in alignment with the identified need to offer students maximum flexibility of choice when selecting courses to take (McCollum, 2007), to improve on measurability and reproducibility of solutions to timetabling problems (Schaerf & Di Gaspero, 2007), to consider the inefficiencies embedded in the input constraints that are provided to the sophisticated optimization algorithms that the research community has developed, and to leverage the existing corpus of knowledge in the field and at the institutional levels (De Causmaecker & Berghe, 2012).

There are initial advances on the referred directions that based on detailed knowledge of an institution's academic programs account for schedule conflicts (Zeising & Jablonski, 2012); that use that knowledge to help transform the curriculum model into the enrollment model (Müller & Rudová, 2012); and that survey faculty on courses that they consider should be offered in non-conflicting schedules (Wehrer & Yellen, 2013).

2 Proposed Approach

Large universities normally have in the order of tens of thousands of students registered in hundreds of programs across schools and departments that offer from hundreds to thousands of sections per term, frequently with multiple sections of the same course offered at different schedules. A common approach to enrollments in U.S. universities is for academic units (i.e. schools and departments) to prepare and offer their class schedules in an independent fashion with little or no coordination across units. Students then enroll in the offered sections across academic units considering the requirements of their programs of study, interests and offerings in the schedule of classes. Students are frequently enrolled in multiple majors and do not advance in curriculum-like synchronization with their peers.

If we are to explicitly consider the enrollment patterns across courses and schools in support of better informed timetabling activities, then it is necessary to: *First*, identify all the unique frequent combination of courses that students take or are able to take per term as well as those combinations that have not been possible due to schedule conflicts but that could be of interest. This is done using enrollment data from recent previous terms. *Second*, identify the course or combinations of courses of interest per term, which are those with section offerings that limit the enrollment options for students. *Third*, enable analyses of data across terms to identify which offerings appear to be limiting the enrollment options over several terms. *Fourth*, develop components to support an interactive interface that enables scheduling authorities to explore and visualize enrollment patterns and if required work on a collaborative fashion to produce better class schedules.

The exhaustive identification of unique course combinations that students take during a term is performed by modeling the problem as an association rules analysis (Agrawal, Imieliński, & Swami, 1993; Agrawal & Srikant, 1994; Krajca, Outrata, & Vychodil, 2011; Srikant, Vu, & Agrawal, 1997). As summary, a transaction is defined as the group of courses that an individual student takes during a term (e.g. MATH 0200, ENGLISH 0220 and ECONOMICS 0100 where the label refers to the academic subject and the number to the catalog/level of the course). Each course is an item in the transaction and the group of courses in the transaction is a course set. Having the data modeled in the described form enables the direct use of association rules algorithms implemented in multiple software packages.

Among others, the output of the association rules analysis indicates how many students enrolled in each course set. As an illustration, one course set could be MATH 0200, ENGLISH 0220 and ECONOMICS 0100 with 100 students. In the next step it is then necessary to identify if 100 students enrolled in that particular course set because no more students were interested (i.e. there is capacity left) or because an enrollment limit was reached due to schedule conflicts or number of seats offered. Those are course sets of interest as they include courses with sections that potentially limit the enrollment options for students. To identify them, it is necessary to analyze the data set at the level of individual sections and schedules.

A backtracking algorithm and methodologies are proposed to identify the course sets of interest considering actual enrollments, enrollment limits and schedules at the level of individual sections on each course set. A de-normalized relational schema is proposed to support the operations of the referred algorithm, longitudinal analyses of historical enrollment data. Other architectural components that will facilitate a collaborative approach to timetabling will be discussed in a separate article.

Prototypes of the referred relational schema and algorithm have been developed and are being tested and refined using actual undergraduate enrollment data from five recent fall terms at the University of Pittsburgh (Pitt) main campus. This campus enrolls approximately 19,000 undergraduate students in 11 of its 17 academic units; they take about 90,000 seats in 3,000 sections each term. Approximately 80% of the students enroll in three or more different subjects across different departments and schools with 56% taking classes in four or five subjects. The processing of enrollment data for five recent fall terms renders data sets with an average of 45,000 closed course sets per term. The reason for the relatively large number of closed course sets is that the association rules algorithms consider the course sets that students take and subsets that are frequent.

As illustration, Table 1 below shows a sample of seven course sets obtained using preliminary results from the components referred above on Pitt's undergraduate enrollments during a recent fall term. The results enable the identification of four course sets of interest in the sample group that appear to be limiting the enrollment options for students. Students enroll in sections of their preference as available in the offered schedule. After the enrollment period is closed, two of the course sets have subjects with sections that reached or exceeded the enrollment capacity offered and two of them have schedule conflicts that limit further enrollments in the course set.

Sample Course Sets	Enrolled in Set	Seats Left in Set	Interest
ARTSC_CHEM_0960, ARTSC_MATH_0220, ENGR_ENGR_0081	243	0	Oversubscribed: ENGR 0081
ARTSC_CHEM_0960, ARTSC_MATH_0220, ENGR_ENGR_0011	241	26	
ARTSC_ECON_0100, ARTSC_ENGCMP_0200, ARTSC_MATH_0120, CBA_BUS_0010	59	8	
ARTSC_ECON_0100, ARTSC_ENGCMP_0200, ARTSC_MUSIC_0711	12	0	Oversubscribed: MUSIC 0711
ARTSC_ITAL_0001, ARTSC_MATH_0120, CBA_BUS_0010	9	8	
ENGR_CHE_0100, ENGR_CHE_0101, ENGR_ENGR_0020	6	0	Conflict: ENGR 0020
CBA_BUSACC_0040, CBA_BUSHRM_1050, CBA_BUSMKT_1441	4	0	Conflict: CBA_BUSACC_0040

Notes:

Course set labels indicate a course School, Subject Code and Catalog Number

ARTSC	The Kenneth Dietrich School of Arts and Sciences		
CBA	College of Business Administration		
ENGR	The Swanson School of Engineering		
BUS…	Business (ACC: Accounting; HRM: Human Resources Management; MKT: Marketing)		
ENGR…	Engineering (CHE: Chemical Eng)		
CHEM	Chemistry	ITAL	Italian
ECON	Economics	MATH	Mathematics
ENGCMP	English Composition	MUSIC	Music

Table 1: Sample of course sets from actual undergraduate enrollments during a recent fall term at the University of Pittsburgh

Table 2 below illustrates the details on schedules and enrollments on all sections for one of the course sets listed in Table 1 (ENGR_CHE_0100, ENGR_CHE_0101, ENGR_ENGR_0020) after the enrollment period has ended. Even though there were seats left in sections of each of the courses, it appears that it was ultimately a schedule conflict that prevented more than six students from enrolling in this course set during the analyzed term.

Course #	School	Subject Code	Catalog #	Class #	Days	Start Time	Stop Time	Enrollment CAP	Enrolled	Seats Left	Seats Left Catalog
1	ENGR	ENGR	0020	14327	M W	16:00	17:15	70	71	-1	10
1	ENGR	ENGR	0020	14443	T H	9:30	10:45	70	70	0	10
1	ENGR	ENGR	0020	14566	T H	9:30	10:45	70	59	11	10
2	ENGR	CHE	0100	14484	M W F	8:00	9:50	65	56	9	18
2	ENGR	CHE	0100	23971	M W F	8:00	9:50	65	56	9	18
3	ENGR	CHE	0101	14485	H	8:00	9:50	65	53	12	25
3	ENGR	CHE	0101	23963	T	8:00	9:50	65	52	13	25

Table 2: Schedules and enrollments on individual sections offered in a selected course set

Currently, work is advancing on: *First*, algorithm improvements to reduce processing time. *Second*, there is ongoing development of analyses and metrics across several terms (i.e. longitudinal analyses) to provide better and more informative identification of course sets of interest. That is, those course sets that appear to limit enrollments across multiple terms. *Third*, components are being extended to enable the identification of negative association rules. This entails the identification of n-tuples of courses that cannot be taken together due to schedule conflicts. Those *n-tuples* do not show up in the association rules analysis as students cannot take those course sets. *Fourth*, development of a graph/network based visualization to facilitate understanding of enrollment patterns and identification of courses that would benefit from collaborative scheduling efforts.

References

Agrawal, R., Imieliński, T., & Swami, A. (1993). *Mining association rules between sets of items in large databases.* Paper presented at the ACM SIGMOD Record.

Agrawal, R., & Srikant, R. (1994). *Fast algorithms for mining association rules.* Paper presented at the Proc. 20th Int. Conf. Very Large Data Bases, VLDB.

De Causmaecker, P., & Berghe, G. V. (2012). Towards a reference model for timetabling and rostering. *Annals of Operations Research, 194*(1), 167-176.

Krajca, P., Outrata, J., & Vychodil, V. (2011). *Using frequent closed itemsets for data dimensionality reduction.* Paper presented at the Data Mining (ICDM), 2011 IEEE 11th International Conference on.

McCollum, B. (2007). A perspective on bridging the gap between theory and practice in university timetabling *Practice and Theory of Automated Timetabling VI* (pp. 3-23): Springer.

Müller, T., & Rudová, H. (2012). *Real-life curriculum-based timetabling.* Paper presented at the Dag Kjenstad, Atle Riise, Tomas Eric Nordlander, Barry McCollum and Edmund Burke. Proccedings of the 9th International Conference on the Practice and Theory of Automated Timetabling. Son, Norway: SINTEF.

Schaerf, A., & Di Gaspero, L. (2007). Measurability and reproducibility in university timetabling research: discussion and proposals *Practice and Theory of Automated Timetabling VI* (pp. 40-49): Springer.

Srikant, R., Vu, Q., & Agrawal, R. (1997). *Mining association rules with item constraints.* Paper presented at the KDD.

Wehrer, A., & Yellen, J. (2013). The design and implementation of an interactive course-timetabling system. *Annals of Operations Research*, 1-19.

Zeising, M., & Jablonski, S. (2012). *A Generic Approach to Interactive University Timetabling.* Paper presented at the ACHI 2012, The Fifth International Conference on Advances in Computer-Human Interactions.