A Bayesian Approach to University Venue Allocation

Mmasehume Raphiri
School of Computer Science and Applied Mathematics
The University of the Witwatersrand
Johannesburg, South Africa
mmasehume.raphiri1@students.wits.ac.za

Ritesh Ajoodha
School of Computer Science and Applied Mathematics
The University of the Witwatersrand, Johannesburg
Johannesburg, South Africa
ritesh.ajoodha@wits.ac.za

***Abstract*—The University Venue Allocation is a prominent and ever-increasingly NP-hard assignment problem in which students are assigned to available venues. A significant corpus of literature has explored various research directions to solve this problem. However, the uncertainty of conflicting constraints when solving the problem has not been probed extensively. This paper uses a Bayesian network (BN) modelling approach to address both uncertainty and causality in the venue allocation problem, and optimise the assignment of venues. Factors influencing venue allocation and the causal relationships between them were identified. Thereafter, a Bayesian Network structure of the problem is constructed. Data is then synthesized from the baseline model using BN forward propagation. Structure learning is carried out using a local search procedure (greedy hill-climbing) and Bayesian Parameter Estimation (BPE) is then used to learn the parameters of the model from the dataset. The model proposed in this paper is capable of resolving some of the most important issues with current venue allocation approaches by reducing uncertainty in expert evaluations and modelling the causality between various sources of schedule uncertainty.**

***Index Terms*—Bayesian networks, University venue allocation, Uncertainty**

I. INTRODUCTION

The University venue allocation problem, which is a subproblem of the University Course Timetabling Problem (UCTP), is a type of resource-constrained scheduling problem which involves the allocation of courses to venues, and possibly other resources subject to some constraints [1]. The constraints of any timetabling problem are classified as either soft or hard constraints, where, the former may be broken as long as the penalty cost is maintained low and the latter must be satisfied for the solution to be feasible [2].

Effective course scheduling at universities is critical for optimal resource usage. Traditionally, the problem is tackled manually by trial and error, with no certainty of a valid and effective timetable solution [3]. This may result in difficulties where the unavailability of a venue impacts the entire learning process, necessitating a manual rearrangement of the timetable. Even if a valid timetable solution is established, it is likely to raise the expense of the rescheduling procedure, and more crucially, it is likely to lose out on more effective options. As a result, these uncertainties have remained the primary motivation for scientific research on the problem.

The need for optimal teaching conditions in universities is another reason it is necessary to solve this problem. Too many students in small venues result in a decrease in the quality of learning and students who require more support are often disregarded. Classes being allocated to venues that are too far away from each other leads to a needless waste of time since students and staff must walk to these remote venues.

Due to the prominence of the UCTP, numerous attempts have been made to either solve or provide more optimal solutions to the problem. These include a metaheuristic approach based on Simulated Annealing proposed by [4] in post-enrollment course timetabling. The proposed approach was evaluated on a benchmark dataset, which demonstrated that it was properly developed and modified. The proposed solution performed well on all datasets and produced new optimal solutions for the majority of the instances. [5] proposed the use of two-staged Graph Coloring to find feasible solutions to the UCTP. The Least Saturation Degree First (LSDF) method was applied in the first stage to discover feasible solutions. The solution quality was improved in the second stage by employing operators based on column permutation. Socha benchmark datasets were used to evaluate the algorithm and it produced encouraging results.

Venue allocation inevitably involves uncertainty. One or more of the following causes this uncertainty:

- Causal relationships between constraints.
- The number of available and required venues.
- Occurrence of conflicting constraints.
- Using data that is subjective rather than objective.

Expert evaluations are extremely subjective, which makes quantifying uncertainty challenging. Despite the extensive research and variety of approaches to solving the venue allocation problem, this uncertainty is not studied in the current venue allocation literature.

The most well-established approach to overcoming uncertainty in these circumstances is the Bayesian method [6]. The Bayesian approach is extended to situations involving intricate causal relationships through the use of Bayesian networks. This paper proposes a Bayesian network approach to solve the venue allocation problem and optimise the assignment of venues. The intuition of our approach is to generate optimal venue assignments by learning the factors that influence uncertainty. We firstly define all the factors that influence venue allocation and use Bayesian structure learning to construct an allocation model, and thereafter, synthesise large amounts of data to train the model and get the conditional probability tables. Casual inference was conducted to check the robustness of the proposed approach.

This paper provides the following contributions:

1) this paper provides a Bayesian approach to solving the venue allocation problem;
2) we address the gap in literature of modelling uncertainty;
3) we provide a framework for the development of new venue allocation solutions that are better informed and hence, are more useful and reliable

(b) we provide a framework for the new

The rest of this paper is structured as follows: Section II reviews some of the current work in timetabling; Section II presents the research question as well as the aim of the research, a specification of our research methodology is then provided in Section IV, which includes a description of the data and testing of the model; and then, Section VI and Section V review the results and conclusion of this work respectively.

II. Related Work

Timetabling problems have been solved using a variety of algorithms. The earliest algorithmic approaches to the timetabling problem were derived from graph colouring heuristics. As seen in the work done by [7], these algorithms perform admirably in small instances of timetabling problems, but not in large ones. [8] noted that the drawback of using graph colouring algorithms to solve timetabling problems is that they require the inclusion of non-academic limitations in problem formulation, which makes the solution hard to implement.

Recently, metaheuristic approaches have been extensively used to solve this problem and optimize timetable generation. Ant Colony Optimisation [[13]; [9]; [4]; [12]] and Genetic Algorithms [[10]; [11]] are some of the most well-known metaheuristics that have gotten the greatest attention from researchers and have been effectively applied to a wide range of optimization issues. Burke and Petrovic [HERE] assert that metaheuristics can perform well in complex issues; however, they can take a long time to solve timetabling problems.

Timetabling problems have been reported to be solved by various combinations of population- and local-area-based algorithms in the literature []. Due to the fact that population-based algorithms are more focused on exploration than exploitation, the quality of a solution generated by a population-based algorithm may not be superior to that of local-area-based algorithms.

III. Research Question

In this paper, the following assumptions were made; students have already been assigned to courses, students and lecturers have no class preferences and time is a flexible constraint. These assumptions narrow down the scope of the problem and allow us to model it as just a venue allocation problem. The first one helps constrain the boundaries of our study domain to just university course timetabling compared to subject allocation. Then, the second assumption prevents us from modelling our problem as just a constraint satisfaction problem. Finally, the last one allows us to remove the added complexity that comes with assigning lectures into timeslots.

To achieve the purpose of this research, given the gap in literature and the above assumptions, the following research question will be explored: Can a Bayesian network-based approach be used to solve the venue allocation problem?

To answer this question, firstly we will define all the factors that influence venue allocation and use Bayesian structure learning to construct an allocation model. Secondly, we synthesise large amounts of data to train the model and get the conditional probability tables. Lastly, we will evaluate the score of the learned model and do some casual inference to check the model's robustness.

A set of constraints must be satisfied by the final solution. The constraints are classified as either soft or hard constraints. The solution is feasible if it satisfies all the hard constraints. Hard constraints are:

1) Students can only attend one class at a time.
2) A venue can accommodate only one class at a time.
3) The number of seats in a venue should be sufficient for all students in a class.

Soft constraints may be broken as long as the penalty cost is maintained low. Soft constraints are:

1) Venues should not be far away from each other.
2) The number of empty seats should be minimized.

A. *Purpose Statement*

It is evident from the literature that there is a huge gap in the university allocation problem in terms of quantifying uncertainty. A step towards filling this gap would be to address some courses of this uncertainty through: (1) modelling the causality between various sources of venue allocation uncertainty, (2) evaluating the effects of conflicting constraints and (3) using large amounts of data to learn the parameters of the model and predict venue assignments. This approach would allow universities to efficiently allocate the best venue to each class resulting in less unnecessary cost overheads.

IV. Research Methodology

A. *Data*

Since the UCTP, in general, is non-specific and different for each institution, the data used for learning and testing is entirely synthetic. BN forward propagation was used to synthesise data. The baseline Bayesian network model was curated to model a situation that could occur in a university context so the data is not purely random.

The following steps are followed to generate the synthetic data:

1) We designed a BN by creating a casual acyclic graph consisting of vertices and edges.

2) A vertex is created for all the variables and the edges represent the independence assumptions between them.
3) Using the assumptions we inferred conditional probabilities between the variables.
4) We simulate data using Bayesian network forward propagation.

The data consists of a selection of classes to be assigned to appropriate venues. The classes are influenced by the number of students and the type of class they need to attend. The selection of venues is based on the size of the venue, the type of venue and the distance between venues. This is scalable because there is no limit to the number of classes that need to be assigned venues. The data is discrete and only has information about the occurrence of each variable. Since the data is synthetic, there was no need to do any explicit pre-processing.

B. Representation

The data used was sampled from a Bayesian network structure shown in Figure 2. Each node represents a factor influencing venue allocation. The edges indicate that one node directly impacts the other node, while the absence of a directed edge indicates that the nodes are independent of one another. Bayes' Theorem, which is described in terms of conditional probability, is used to establish this causal relationship:

Where:

A : is the hypothesis.
B : is the evidence.
$P(A)$: is the likelihood of the hypothesis being true before the evidence is present.
$P(B)$: is the likelihood of observing the evidence.
$P(B|A)$: is the likelihood of observing the evidence if the hypothesis is true.

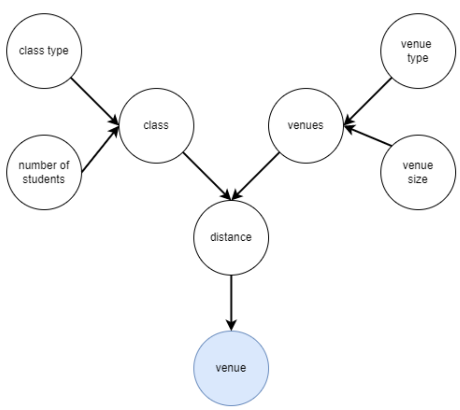

Fig. 1. The Bayesian Network Structure for Venue Allocation

C. Learning

1) Structure Learning: The cornerstone of Bayesian network learning is structure learning, and effective structure learning is the key to building the optimal network structure since the network structure and data set can be used to determine the parameters. Score-based structure learning methods attempt to identify the best structure for a given dataset by maximizing a score while traversing the space of potential structures. Using a data fusion approach, we obtained a Bayesian network structure based on expert experience and machine learning of data.

TABLE I
DATA REPRESENTATION

Feature	Description	Example
Number of students	The number of students in the class.	250
Class	Name of the class (course).	COMS1018A
Class Type	Classifies the classes into three class types: Tut, Lecture, and Lab.	Tutorial
Distance	Distance between venue and faculty office.	10m
Venue Size	The number of students a venue can accommodate.	400
Venues	List of all values that satisfy the venue size and venue type criteria.	
Venue Type	Classifies the venue into three venue types: Tut, Lecture, Lab.	Tutorial
Venue	Name of the venue.	SHBB

We used Hill-Climb Search as the search strategy and we benchmarked scoring functions, K2 and Bayesian Information Criterion (BIC), to select the best scoring function to measure the fit between the structure and the data. Benchmarking results for BIC score were significantly better than the BD score thus we selected BIC score.

2) Parameter Learning: We trained the BN classifiers to identify the ideal Bayesian structures by estimating a parameter set of joint probability distributions that best reflect supplied data set with labelled examples. Maximum Likelihood Estimation and Bayesian Parameter Estimation were explored to learn the parameters of the model and the results we evaluated.

Maximum Likelihood Estimation (MLE) estimated the parameters of the network, using the data set. To find the MLE we utilized the relative frequency, with which the variable states have occurred and filled the conditional probability tables (CPD) in such a way, that *P(data|model)* is maximal. The problem with MLE was that it overfit data and when the observed data was not representing the underlying distribution the MLEs were extremely inaccurate. As a result, MLE was extremely weak and unstable for learning network parameters. To mitigate MLE's overfitting we computed Bayesian Parameter Estimation (BPE).

To find the BPE we started with previously existing CPDs that describe our ideas about the variables prior to data observation. The priors were then updated using state counts from the observed data. We explored two priors, the K2 prior which simply added one to every state count and Bayesian Dirichlet equivalent uniform (BDeu) prior which finds the Maximum a posteriori (MAP) structure, ensuring likelihood equivalence and equivalent sample size. BDeu prior was a more sensible choice of prior.

V. Results

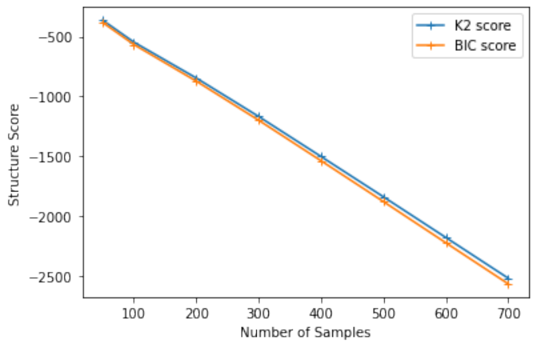

Fig. 2. Structure Score

TABLE II
Log likelihood Score Averaged Over 10 Runs

Number of Samples	Baseline	BPE with BDeu
0	0	0
50	-396.7086	-325.9922
100	-576.8100	-498.1301
200	-886.0752	-785.967
300	-1213.8000	-1106.6489
400	-1529.4773	-1414.5612
500	-1881.4774	-1756.5213
600	-2233.7893	-2098.3670
700	-2571.5825	-2437.0780

VI. Conclusion

Venue assignments are critical to the smooth running of universities. Venue allocation under uncertainty has not been extensively studied in literature. This paper proposed a Bayesian network to construct a model for the problem and try to generate optimal venue assignments.

References

[1] Chen, M., Tang, X., Song, T., Wu, C., Liu, S., & Peng, X. (2020). A tabu search algorithm with controlled randomization for constructing feasible university course timetables. Computers & Operations Research, 123, 105007.

[2] A survey of metaheuristic-based techniques for university timetabling problems. OR spectrum, 30(1), 167–190.

[3] Basir, N., Ismail, W., & Norwawi, N. M. (2013). A simulated annealing for tahmidi course timetabling. Procedia Technology, 11, 437–445.

[4] Ceschia, S., Di Gaspero, L., & Schaerf, A. (2012). Design, engineering, and experimental analysis of a simulated annealing approach to the post-enrolment course timetabling problem. Computers & Operations Research, 39(7), 1615–1624.

[5] Budiono, T. A., & Wong, K. W. (2012). A pure graph coloring constructive heuristic in timetabling. In 2012 international conference on computer & information science (iccis) (Vol. 1, pp. 307–312).

[6] Efron, B. (2005). Bayesians, frequentists, and scientists. Journal of the American Statistical Association, 100 (469), 1–5.

[7] Welsh, D. J., & Powell, M. B. (1967). An upper bound for the chromatic number of a graph and its application to timetabling problems. The Computer Journal, 10 (1), 85–86.

[8] Yu, E., & Sung, K.-S. (2002). A genetic algorithm for a university weekly courses timetabling problem. International transactions in operational research, 9 (6), 703–717.

[9] Cauvery, N. (2011). Timetable scheduling using graph coloring. International Journal of P2P Network Trends and Technology, 1(2), 57–62.

[10] Abdullah, S., Burke, E. K., & McCollum, B. (2007). A hybrid evolutionary approach to the university course timetabling problem. In 2007 ieee congress on evolutionary computation (pp. 1764–1768).

[11] Khonggamnerd, P., & Innet, S. (2009). On improvement of effectiveness in automatic university timetabling arrangement with applied genetic algorithm. In 2009 fourth international conference on computer sciences and convergence information technology (pp. 1266–1270).

[12] Nothegger, C., Mayer, A., Chwatal, A., & Raidl, G. R. (2012). Solving the post enrolment course timetabling problem by ant colony optimization. Annals of Operations Research, 194(1), 325–339.

[13] Socha, K., Knowles, J., & Sampels, M. (2002). A max-min ant system for the university course timetabling problem. In International workshop on ant algorithms (pp. 1–13).