Scientific Research

Enhanced Greedy Optimization Algorithm with Data Warehousing for Automated Nurse Scheduling System

R. M. Kapila Taranga Ratnayaka[1], Zhong-Jun Wang[1], Satish Anamalamudi[2], Siyue Cheng[3]

[1]Department of Statistics, Science College, Wuhan University of Technology, Wuhan, China
[2]School of Information Engineering, Wuhan University of Technology, Wuhan, China
[3]College of Engineering, Colorado State University, Fort Collins, USA
Email: kapila.tr@gmail.com, kapilar@sab.ac.lk

ABSTRACT

Current Nurse scheduling process has many challenges like work plan creation and working hour allocation for employees at specific planning horizon. Hospitals in most of the developing countries use manual methods to create nurse scheduling systems. With current existing manual nurse scheduling systems, most of the hospitals especially in developing countries don't have efficient work plan allocation. Moreover, patients need nursing care throughout the day. Hence, current manual nurse scheduling approach with simple statistical functions is not efficient especially for highly populated countries. Our proposed automated nurse scheduling approach has carried out in two stages. Firstly, we propose an efficient data warehouse system based on online analytical method for hospital information system. Subsequently, Enhanced Greedy Optimization algorithm is implemented to optimize the nurse roster and compared with other optimization algorithms (Simulated Annealing and Genetic Algorithm). Experimental results (MYSQL, JAVA, OLAP) with proposed optimization algorithm outperforms compared with existing optimization solutions.

Keywords: Java; MYSQL; Date Warehouse; OLAP

1. Introduction

Nurse scheduling process is very crucial task for managing nurse duty schedules in hospital system. It is a process of creating a balanced work schedule for nurses. Currently, even in most of the reputed hospitals, nurses spend additional time for creating nurse schedule manually which is very complex and not effective. Since people need healthcare throughout 24 hours, it is important to design efficient automated nurse scheduling software to manage nurse duty activities.

The feasible scheduling system has main impact on the quality of the health care and their budgets. Hence, hospital management system has to concern many constraints at the time of implementing automated nurse schedule [1,2]. Various necessities of patients, different type of qualifications, experiences and specializations of nurses, employers and employee requirements, unpredictable incidence and absenteeism and other factors make the problem more complicated.

In our proposed work, main focus is on the information based database model for nurse management system especially for highly populated countries [3,4]. Usually, nurse in charge of each ward has responsibility to maintain records in their respective ward. Based on the prepared schedule, nurses will be assigned for working shifts with consideration of time, requirement, along with experience and nurse skills. In this scenario, manual system for maintaining records is not effective especially at the time of holidays and busy working hours where most of the nurses might be on leave [4]. With advance in IT technology and efficient optimization algorithms it is not so difficult to design efficient automated nurse scheduling software with best optimization algorithm. Moreover, it is flexible, reliable and more efficient compared with manual nurse scheduling.

In our approach, data warehouse based on medical information system is used to store, retrieve and manage the nurse information. Considering general and individual factors, efficient flexible nurse management warehouse based online analytical process (OLAP) has been proposed. The rest of the paper is organized as follows. Section 2 explains about brief overview of existing solutions with pros and cons. Section 3 explains about proposed work with enhanced greedy optimization algorithm. Section 4 explains about experimental results and Section 5 ends up with conclusion and future work.

2. Related Works

Existing research works has been proposed diverse models and methodologies to develop automated nurse sched-

uling problems. Most of the current proposed solutions either make use of random based optimization algorithms which won't be efficient or applicable only for fully automated nurse scheduling problem. Li Ping [5], Wu Tao, Chen Mu, Zhou Bin and Xu Wei-guo has done much research work related to Medical informatics in 2011 and had deep discussions about the role of data warehouse management system to handle hospital and nurse management information [6]. Multidimensional analysis techniques under different angles were used to extract the required data and information. Michael Silver [7] and his group used data mining techniques for data warehouse and published their findings under the title "Case study: How to apply data mining techniques in a Healthcare Data warehouse". This approach has been implemented successfully in many of the American hospitals [8]. Two numerous data mining techniques called; patient rule introduction method (PRMI) and weighted item sets (WLS) were used to analyse large quantities of data. In 1998, Peter Villiers and his team worked together to apply data mining techniques for solving clinical data warehouse functionality and proposed Flexible clinical data mining system (CDMS) using SAS statistical software [9,10]. In addition, research is carried out in two stages. In first stage, controlled environment were provided for CDMS access based systems and transformed it into analytical clinical data. In the later stage, operations were tested with the row data operations with same data. Peter Villiers proposes genomic based data for further performance enhancements.

In 2008, S. Kundu and M. Mahato described the use of Genetic Algorithm (GA) for solving NSP. They used two different models, Simulated Annealing and Genetic Algorithm to solve this problem. Compare nurse performance at different levels. They have considered soft and hard constraints [1].

End of 2009 K. Jaumard reported a method to solve the nurse roster problem using column generation. There sub problem was formulated as a shortest path problem with resource constraints, where each possible shift was represented by a node and It was solved by using a two-stage algorithm.

2.1. Problem Definition

Organizations that operate continuously can divide their daily works into shifts. In such scenario, scheduling approach to assign work schedule to each worker, which involves building a timetable for specified period is required. In such scenario, efficient evolutionary based algorithms should be implemented along with data warehouse to enhance the performance of automated nurse scheduling approach. Most of the current research works [11] has been proposed with either random based opti-

mization algorithms or local optimization algorithms which cannot withstand for difficult scenarios. Moreover, it may lead to local optimization which causes severe performance degradation. In addition, there are cases where nurses may change their present shift, while other nurses are scheduled around this pre-shift. In this case, hospital management with manual nurse scheduling can face many difficulties to assign works for nurses in different wards to the shifts according to the requirements. This is one important and challenging consideration which has not been proposed in any of the current research works. In our proposed work, a module is designed which can dynamically mange the shift change of nurse whenever someone needs to change their shift time. This needs to maintain a list of nurse information for those who are going to work in next given schedule and those who are on leave. Whenever, some wants change the shift, our proposed module will shows the list of names which is easy to adjust with other nurses. Moreover, each and every staff member should be equally allocated for the night shifts, off days in weekends and public holidays. Also different type of qualifications, skills and experiences, different demands of patients, unpredictable absenteeism and other factors make the problem complicated. The objectives in this problem are multiple and complicated. So, our main target is to find high quality feasible schedule and resource assignments under the labour contract rules and satisfying employees as well as employer's requirements and constraints. Several requirements and constraints were identified during the requirement analysis. Easy to understand, it is divided in to three categories.

- Manpower demand and working hour constraints

Working experiences and rank, staff group, skills, gender, total number of working hours per month including OT and other special qualifications were based on this category.

- Shift distribution and sequence pattern constraints

This category include the number of working hours and pattern of shifts to be assign on consecutive days for each and every staff member within the month [12].

- Maximum and minimum shift constraints

The number of shift per month assigned to each nurse must be within the limits of legal regulations.

In general, head nurse has a responsibility to construct nurse roster and should be published before next month in current manual nurse scheduling process. These traditional database systems are not well suitable to deal with statistical techniques and quick decisions. Moreover, it makes so redundant and surplus works [13]. In this paper, we focus to find a feasible solution for nurse scheduling system considering mentioned constraints and regulations based on data warehouse with online analytical processing (OLAP) techniques. Java, PHP and MYSQL

were used to handle programming language and database management systems.

3. Proposed Work

Nurse scheduling system represents the important administrative activity in real world modern hospitals. Major task is to identify the main areas, main working categories and allocation of recourses in efficient way. Based on requirement analysis, five main categories were identified. They were administrators (employer), Head nurse and nurses (employee), patients (customers) and other staff members.

3.1. OLAP in the Multidimensional Data Model

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

Data storage, retrieval and management are one of the crucial factors for managing efficient automated nurse scheduling software. Hence, data warehouse and OLAP tools based on a multidimensional model has been used in transformation process. Nurse Management system along with hospital information management system includes monthly working schedule (nurse roster), minimum and maximum monthly workload, Special duties with OT periods, working preferences, requirements, skill levels, categories, previous working records, workshops, educational information and personal information. Moreover, it should be linked with management of patient information, management of medicine information and hospital business information systems. In multidimensional model, data are organized into multiple dimensions, and each dimension contains multiple level of abstraction defined by concept hierarchies. Nurse scheduling system cube contains three dimensions such as; Department, Ward and seniority. Also data warehouse often adopt three-tier architecture: bottom tier (data warehouse server), middle tier (OLAP server) and top tier (front-end tools).

- Bottom tier layer: The bottom tier is a warehouse data base server that is almost a relational database system. It is the first layer which stores pure information. This tier also connects to the metadata repository, which stores information about data warehouse and its content.
- Middle tier (OLAP server): The middle tier is an OLAP server, which handles the request of the client to complete some logic task. Normally OLAP servers present business with multidimensional data from the data warehouse or data mart. However, the physical architecture and implementation of must consider

data storage issues.

Top tier: This is a user interface layer, which contains query, reporting analysis and data mining tools. The architecture and functional overview of data warehouse is shown in **Figure 1**.

3.2. Design of a Data warehouse: Roster Analysis Framework

Proposed solution is a web based automated system for nurses scheduling. Hence, one can update and view their records in online. The data warehouse for the hospital is a large-scale database. As a first step, Administrator in the Hospital management system is a main user. The proposed system mainly divided in to two parts called; hospital management system and clinical information system. **Figure 1** clearly explains about OLAP design model for Nurse scheduling problem.

The design mode of the OLAP model based data warehouse is based on **Figure 1**. Hospital management system includes Management information unit of employees and Pharmacy information database. Clinical information system includes Accident and Emergency Unit, Patient information unit (OPD and home), Surgery and Clinical information, Maternity unit, Dental unit and electronic medical information database. For reliable handing, administrator appointed department heads for each and every department. Department heads have a responsibility to maintain their departments' accounts. So department heads appointed head nurses for every ward.

Head nurse is a main user in the ward and she creates new accounts for nurses in her ward. Hence, all the nurses must register the system and need to create their own accounts. Once user gets activated their access, system provides a facility to update their details, apply leaves, apply and conform OT, view their time tables for nurses. So, nurses must send their leave details through their accounts before creation the roster. In this proposed system, Modified Greedy algorithm is used to create feasible high quality roster. Based on the output of optimization algorithm, scheduling tasks will be activated and
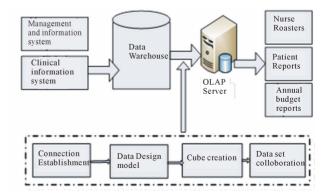

Figure 1. Design overview of NSP in OLAP.

assigned to corresponding users.

3.3. Enhanced Greedy Optimization Algorithm

Greedy Algorithm follows heuristic problem solving with locally optimal choice at each stage with the hope of finding a global optimum. "Traveling salesman problem" is one of the best examples for heuristic based approach. In our approach, Enhanced Greed optimization is used to generate best global optimized value for nurse scheduling.

Proposed Enhanced Greedy Algorithm has five crucial components

- Candidate set should be created.
- Selection function should be defined so that it selects the best candidate to add to the solution.
- Feasibility function should be defined to identify whenever a candidate can be used to contribute to a solution.
- Modified objective function is used to assign a value to local solution, or a partial solution,
- Solution function helps to identify the complete solution of the intended problem.

Proposed Greedy algorithm is divided into two categories namely (1) Degree based Ordering and (2) Grouping. The algorithm for degree based ordering with graph colouring approach is explained in (1).

Based on proposed algorithm, first matrix is created for the nurses and monthly dates where nurse names were taken as rows and days of the month was taken as columns. In matrix, nurse names will be "i" and days of the month will be "j". Hence it explains total number of nurses working in a ward for 31 days of the month. Later, matrix is divided into four equal sub matrices. Firstly, night shifts for Diagonal of the first sub matrix were assigned. Secondly, night shifts were assigned for diagonal of the fourth sub matrix. Later night shifts were assigned for 2nd and 3rd sub matrices.

3.4. Degree Based Ordering for Nurse Scheduling : Graph Colouring Approach

```
{
Vertex (G) ← {x_1,x_2.........x_n}
Colour(G) ←     {c_1, c_2,c_3,c_4.....c_n}
for (i=0; i<n; i++) // where n= number of vertices
{
for each colour vertex u -> n_2(v_i) do    // u is subset of 'v'.
{
TabooColors(color(u)) = v_i
}
Colour (v_i) = min{c : TabooColors(c) !=v_i}
loop: Let C_i  be the first colour in C.
For each j with i<j and x_j adjacent to x_i in G
```

Set $C_j = C_j - \{ C_{i\}}$ *//where* x_j *will not have same colour as* x_i
Change i to i+1 and if i+1 ≤ n,
Return to "loop".
}   (1)

Optimization of Nurse scheduling with grouping scenario is explained in (2).

3.5. Greedy Algorithm for Grouping

```
{

Create variable start Nurse(SN), Start day(SD), end nurse (EN) & end day(ED);

Nurses working in ward=r;
Working days =q;
Create 2 way integer array {NURSESSHIFTS }

 // Rows ← Day & column ← Nurse identity
If(nurse shift== True)
{
Shift←1;
}
else
shift←0;
Group ← nurse number with group identity
One way group ← Array {Category}
Ward number ←x;

Assign (r/2)^{th}(Q/2)^{th} position of nurse shifts
Send (SN=r/2), (SD=q/2),(ED=r)
(ED=q) to scheduler Filling(SF)
Initiate (SN=0), (SD=q/2),(EN=r/2) and (ED=q) to SF

Send (SN=r/2), (SD=0),(EN=r) and (ED=q/2) to SF
Send (SN=0), (SD=0),(EN=r/2) and (ED=q/2) to SF

}                                                                              (2)
```

Optimized solution of the nurse scheduling can be generated with set of feasible solutions and optimization function.

At each generation, decision should be best with respect to time.

On the other hand, simulated annealing and Genetic algorithm is used to compare the results with proposed enhanced greedy optimization algorithm. Comparison with other intelligent optimization is made in terms of start time, stop time and cost function value which is briefly explained in below.

4. Experimental Results

OLAP is used as a data ware house to store, retrieve and manage the nurse information. MYSQL is used to access

information from data warehouse with Java frontend. Firstly, proposed Greedy optimization algorithm is implemented in java. Detailed nurse information is created and accessed through MYSQL. Optimization with predefined cost function and set of feasible solutions are used to determine the best optimized value.

Optimization is done in terms of 7, 14, 21 and 30 days. Problem size is considered as 500 and compared the proposed greedy solutions with simulated annealing and Genetic algorithm. **Table 1** explains about the comparison of proposed results with simulated annealing and GA with respect to time in seconds.

Firstly, average time taken for greedy is 0.67 sec whereas for SA it is 0.77 and GA it is 2.5 seconds. For, 14 days greedy has taken 1.34 sec whereas SA has taken 2.85 and 7.21 for GA. Similarly, for 21 day schedule, 3.21 is for greedy, 3.48 for SA and 8.26 for GA. Compared with 7, 14, 21 day simulation results, 30 day simulation result is almost same for Greedy, SA and GA. From these, one can say that for few day schedules greedy and simulated annealing performance is very near compared with Genetic optimizations. For long period simulations, all the three optimization algorithms holds good to get optimized nurse schedule results. Hence, it is good to choose greedy or simulated annealing for short period nurse scheduling and can use genetic optimizatio- for long period nurse schedule.

Gnu plot is used for graphical representation of simulation results that were compared from optimization algorithms.

5. Conclusion and Further Works

Nurse scheduling is one of the crucial problem that are facing in many hospitals all around the world. Some con- straints make the problem more complex and complicated. Currently high qualified health personals have been conducting so many researches to find fair and better solution. This research described new information system based on OLAP data warehouse techniques. Moreover, this study opens a new path for hospital information systems to deal with statistical methods such as data ware house and data mining.

New proposed system helps to keep their records foundation for making future decisions. Our proposed work outperforms compared with existing manual system and automated local optimization algorithms.

To fulfill the business requirements, there are few other constraints that have to be added with current proposed solution. In future, enhancements like linking databases for OLAP online system can be done to make it as high quality and reliable solution.

REFERENCES

[1] S. Kundu, M. Mahato, B. Mahanty and S. Acharyya, "Comparative Performance of Simulated Annealing and Genetic Algorithm in Solving Nurse Scheduling Problem," *Proceedings of the International MultiConference of Engineers and Computer Scientists*, Hong Kong, 19-21 March 2008, p. 96.

[2] G. Baskaran, A. Bargiela and R. Qu, "Hierarchical Method for Nurse Rostering Based on Granular Pre-Processing of Constraints," *The* 23*rd EUROPEAN Conference on Modelling and Simulation*, Madrid, 9-12 June 2009, pp. 855-861.

[3] Q. Y. Guo, F. D. Hao, X. L. Duan, X. Q. Xie and W. Liao, "Multi Personal Computer Storage System: Solution of Sea Capacity PACS Storage," *Chinese Medical Journal*, Vol. 116, No. 5, 2003, pp. 650-653.

[4] R. Paul and A. S. M. L. Hoque, "A Storage & Search Efficient Representation of Medical Data," 2010 *International Conference on Bioinformatics and Biomedical Technology*, Chengdu, 16-18 April 2010, pp. 418-422.

[5] P. Li, T. Wu, M. Chen, B. Zhou and W.-G. Xu, "A Study on Building Data Warehouse of Hospital Information System," *Chinese Medical Journal*, Vol. 124, No. 15, 2011, pp. 2372-2377.

[6] P. Villiers, "Clinical Data Warehouse Functionality," SAS Institute Inc., New Caledonia, 1998.

[7] M. Silver, T. Sakuta, H.-C. Su, S. B. Dolins and M. J. Oshea, "Case Study: How to Apply Data Mining Tech- nigues in a Healthcare Data Warehouse," *Journal of Heth- care Information Management*, Vol. 15, No. 2, 2001, pp. 155-164.

[8] A. H. W. Chun, S. H. C. Chan, G. P. S. Lam, F. M. F. Tsang, J. Wong and D. W. M. Yeung, "Nurse Rostering at the Hospital Authority of Hong Kong," *Proceedings of the* 17*th National Conference on Artificial Intelligence and* 12*th Conference on Innovative Applications of Artificial Intelligence*, Austin, 30 July-3 August 2000, pp. 951-956.

Table 1. Various optimization algorithms for nurse scheduling problem.

Days	Count	Algorithm	Solved	Avg. Time (Sec)
		MGA	460	0.67
7	500	SA	440	0.77
		GA	400	2.5
		MGA	470	1.34
14	500	SA	460	2.85
		GA	365	7.21
		MGA	470	3.21
21	500	SA	455	3.48
		GA	430	8.26
		MGA	485	11.18
30	500	SA	325	11.77
		GA	320	11.28

[9] M. F. Wisniewski, P. Kieszkowski, B. M. Zagorski, W. E. Trick, M. Sommers and R. A. Weinstein, "Development of a Clinical Data Warehouse for Hospital Infection Control," *Journal of the American Medical Informatics Association*, Vol. 10, No. 5, 2003, pp. 454-462. doi:10.1197/jamia.M1299

[10] T. B. Pederson and C. S. Jensen, "Research Issues in Clinical Data Warehousing," 10*th International Conference on Scientific and Statistical Database Management*, Capri, 1-3 July 1998, pp. 43-52.

[11] C. A. Goble, R. Stevens, G. Ng, S. Bechhofer, N. W. Paton and P. G. Baker, *et al*., "Transparent Access to Multiple Bioinformatics Information Sources," *IBM Systems Journal*, Vol. 40, No. 2, 2001, pp. 532-551. doi:10.1147/sj.402.0532

[12] B. A. Eckman, C. A. Bennett, J. H. Kaufman and J. W. Tenner, "Varieties of Interoperability in the Transformation of the Health-Care Information Infrastructure," *IBM Systems Journal*, Vol. 46, No. 1, 2007, pp. 19-41. doi:10.1147/sj.461.0019

[13] X. Z. Zhou, S. B. Chen, B. Y. Liu, R. S. Zhang, Y. H. Wang and P. Li, *et al*., "Development of Traditional Chinese Medicine Clinical Data Warehouse for Medical Knowledge Discovery and Decision Support," *Artificial Intelligence in Medicine*, Vol. 48, No. 2-3, pp. 139-152.