MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-315

Improving Hospital-Wide Early Resource Allocation through Machine Learning

Daniel Gartner[a,b] and Rema Padman[a]

[a] The H. John Heinz III College, Carnegie Mellon University, Pittsburgh, PA, USA
[b] TUM School of Management, Technische Universität München, München, Germany

Abstract

The objective of this paper is to evaluate the extent to which early determination of diagnosis-related groups (DRGs) can be used for better allocation of scarce hospital resources. When elective patients seek admission, the true DRG, currently determined only at discharge, is unknown. We approach the problem of early DRG determination in three stages: (1) test how much a Naïve Bayes classifier can improve classification accuracy as compared to a hospital's current approach; (2) develop a statistical program that makes admission and scheduling decisions based on the patients' clincial pathways and scarce hospital resources; and (3) feed the DRG as classified by the Naïve Bayes classifier and the hospitals' baseline approach into the model (which we evaluate in simulation). Our results reveal that the DRG grouper performs poorly in classifying the DRG correctly before admission while the Naïve Bayes approach substantially improves the classification task. The results from the connection of the classification method with the mathematical program also reveal that resource allocation decisions can be more effective and efficient with the hybrid approach.

Keywords:

Artificial Intelligence; Diagnosis-Related Groups; Hospital Costs; Linear Programming.

Introduction

Several recent studies have highlighted the challenges facing many OECD (Organization for Economic Cooperation and Development) countries with rising pressure on healthcare systems due to increasing demand and expenditures for healthcare [1]. As a countermeasure, cost-containment policies such as diagnosis-related groups (DRGs) were introduced [10] and are evaluated continuously [11]. When hospital resources such as beds, diagnostic devices, human resources and operating room availability become scarce, the coordination of patient care and patient flow becomes a challenge. The problem is even more complicated when detailed information about patients who seek hospital admission is captured in an unstructured form or is likely to be incomplete in early stages of care.

DRG-based reimbursement of patient services

Within DRG systems, patients are classified into groups with homogeneous clinical characteristics and resources required during treatment. After the treatment, patients' health insurance pays a fixed amount for treatment depending on the assigned DRG. Within this context, uncertainties in care delivery information play an important role in the effective planning of healthcare processes. In early stages of care, the patients' DRG is unclear, because of potentially missing information and unstructured documentation of patient information using free-text [6]. This introduces a significant challenge to the hospital's operations management when assigning patients to scarce hospital resources while economic objectives such as contribution margin maximization are pursued [5].

The role of hospital information systems to improve hospital-wide patient scheduling decisions

By structuring patient information that is documented when patients seek admission, hospitals can leverage this information to better predict patients' resource requirements. When information about each patient's DRG, resource requirements, and clinical pathway [12] is made available before admission, a hospital can potentially improve the effectiveness and the efficiency of resource allocation decisions.

In this paper, we investigate whether machine learning methods can increase the classification accuracy of an inpatient's DRG as compared to the current approach of using a DRG grouper. In particular, our focus is to classify the patient's DRG before admission in order to improve admission decisions via novel statistical models.

We analyze inpatient data from one year consisting of more than 16,000 records from a 350-bed hospital. Our results show that, in general, machine learning approaches can substantially increase early DRG classification accuracy, especially for elective patients who contact the hospital before admission. Moreover, we demonstrate that machine learning techniques combined with mathematical programming can lead to improved resource allocation decisions.

The remainder of this paper is structured as follows. In the next section, we present two approaches demonstrating how early DRGs can be classified. The first classification approach can be seen as the current approach used by our study site, while the second one uses a Naïve Bayes classifier. We then develop a statistical model for patient admission and scheduling under several constraints, such as precedence constraints between clinical activities and resource constraints. The results section describes the data that we employed in our analysis and presents results of the classification and resource allocation task followed by a discussion. In the final section, we summarize our results and provide directions for further research.

Methods

In the following, we present two DRG classification approaches and a model for combining and planning patient flow with admission decisions.

Early DRG classification techniques

Current approach using a DRG grouper

To determine a DRG at discharge, typically, a simple flowchart-based method is used. The method is implemented in a commercial software called DRG grouper. Figure 1 illustrates the DRG-grouping.

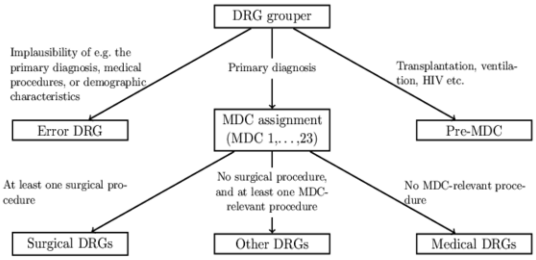

Figure 1 – DRG grouping using a DRG grouper, see Schreyögg et al. [9]

Before the execution of the DRG grouper, parameter values such as the primary and secondary diagnoses, clinical procedures, age, gender, as well as weight in the case of newborns, have to be entered into the software. Diagnoses are coded using the International Statistical Classification of Diseases and Related Health Problems (ICD). The first 3 letters of an ICD code correspond to DRGs. The algorithm first determines one of 23 Major Diagnostic Categories (MDC). These are defined by the primary diagnosis (i.e. the reason for the hospitalization). However, if the primary diagnosis is imprecisely documented, an error DRG will be returned. On the other hand, if the patient has a transplantation, for example, a Pre-MDC (a DRG with high-cost procedures [2]) is returned. After determining the MDC, clinical procedures and co-morbidities lead to the patient's DRG which can be categorized into surgical, medical and other DRGs. Finally, within these categories, the age of the patient, or the weight in the case of newborns, may lead to a different DRG-subtype.

When sufficient information is available, there is valid reason to use a DRG grouper not only at discharge, but at any stage of care in order to obtain real-time information about the patients' DRGs. Therefore, we assume that in every stage of care of a patient and in particular before admission that sufficient information is available to classify the patient's DRG.

Naïve Bayes Classification

Let D be the set of all DRGs, A the set of attributes (e.g. gender, free-text diagnoses) and I be the set of labeled patient instances. The naïve Bayes classifier assumes that all the attributes are conditionally independent given the inpatient's DRG $d \in D$. Under this assumption, the prior probability $p(d)$ of each DRG d is learned from the training data by maximum likelihood estimation, i.e. $p(d)$ is set equal to the proportion of training examples which belong to the class d. Similarly, the conditional likelihood of each value $v_{i,a}$ given instance $i \in I$ and attribute $a \in A$ of each DRG $d \in D$ is learned from the training data by maximum likelihood estimation. This means that $p(v_{i,a} \mid d)$ is set equal to the proportion of training examples of class d which have value $v_{i,a}$ for attribute a in instance i. Afterwards, the classifier assigns the DRG d_i^* to the test instance i by employing Equation (1):

$$d_i^* = \arg\max_{d \in D} \left\{ p(d) \cdot \prod_{a=1}^{|A|} p(v_{i,a} \mid d) \right\} \tag{1}$$

Resource allocation model

In what follows, we present our resource allocation model which is an extension of Gartner and Kolisch's [5] patient flow problem with fixed admission dates. We further extend their model by taking into account admission decisions.

Planning horizon, patients and activities

We have a set of days in which patients can be scheduled and which represents our planning horizon. We have a set of patients who seek admission and can be planned during the planning horizon. The patients have a target admission date and a set of activities to be scheduled within the set of days. For example, an activity can be a computer tomography scan, a surgery or a therapy. Within each patient's set of activities, we have a specific activity which we denote as the discharge activity and which is the final activity of the patient during his/her stay in the hospital.

Time windows, clinical pathways, contribution margin

Each patient's activity has a time window in which it can be scheduled. Once we schedule the discharge activity, the hospital receives a length of stay-dependent contribution margin. The clinical pathway of each patient is represented by precedence relations; each of them can be assigned an integer time lag, which means that the successor activity cannot be scheduled until a certain (waiting) time after the predecessor activity has elapsed. This is necessary to capture recovery times, for example.

Resources, resource capacity and resource requirement

We split hospital resources into a set of day resources and overnight resources as follows: Day resources represent, for example diagnostic devices or the operating room; while overnight resources represent beds since they are typically allocated for at least one night. On the supply-side, each resource has a day-dependent capacity while on the demand-side, each patient's activity has a resource requirement. An activity can require multiple day-resources on the same day, such as when a surgical activity requires the surgery room and the surgeon. Also, an activity does not necessarily require the same amount of time from two distinct resources. For example, a surgical activity may require 100 minutes of surgery room time (including set up time, surgery time, cleaning time, etc.) while the actual time required from the physician may be no more than 80 minutes. Naturally, the documentation task executed by the surgeon follows the surgery and has to be executed before the patient's discharge.

Capacity requirement of overnight resources is typically measured in terms of bed utilization. An overview of all parameters is provided in Table 1 in the Technical Appendix.

Decision variables

Our mathematical model uses binary decision variables to decide whether or not a clinical activity is executed on a day within the activity's time window. Another set of decision variables decides whether the admission of a patient should be avoided. These decision variables inform a decision maker whether or not patients provide the hospital enough of a contribution margin to schedule them along their clinical pathway.

Objective function and constraints

Our mathematical model (its algebraic formulation is shown in the Technical Appendix) maximizes the overall contribution margin of all admitted patients in the hospital by assigning each patient to a discharge date within that patient's discharge time window, see Objective Function (2). Constraints (3) ensure minimum time lags between clinical activities. Note that if we have a time lag of 0 days, the predecessor and successor activity, for example a magnetic resonance tomography and a neurosurgery procedure, respectively, can be performed on the same day. Incorporating the admission variables into the constraints ensures that the constraints are satisfied if the patient is not admitted. Constraints (4) depict the limited capacity of day resources. For each resource and day, capacity demands that all activities performed on that day must not exceed resource capacity. Note that the dimension of resource capacity can be, for example, time or slots in a master surgical schedule or minutes available by a physical therapist. Constraints (5) denote the resource constraints for overnight resources. For each patient, the bed allocation starts with the patient's admission date and the discharge activity releases the bed. Naturally, a patient's resource requirement only sums if he is admitted. Constraints (6) ensure that a patient's activity is scheduled at most one time within the activity's time window. However, once the activity is scheduled, all of the activities corresponding to that patient are scheduled, as required by Constraints (7). Finally, expressions (8)-(9) are the decision variables and their domains.

A bed allocation example

Suppose we have a planning horizon which consists of 7 days and we have two patients to be admitted and scheduled within the planning horizon. Assume that both patients have the actual DRG F21C which means "other surgical procedures with cardiovascular diseases, without complex procedure, without complex skin transplantation and myocutaneous flap surgery of the lower limb admission". The property of the contribution margin function of this DRG reveals that on day 6, it has its maximum. Before that length of stay (LOS), the patient's LOS is below the low LOS trim point. Accordingly, the hospital receives a deductible if the patient is discharged before that day. If the patient's LOS is between the low and high LOS trim point, the contribution margin function is monotonicly decreasing due to LOS-dependent costs for activities such as bed cleaning. We assume that the clinical pathway of the patient requires a minimum time lag of 5 days between surgery and discharge. As a consequence, the time window of the discharge activity has a lower bound for the LOS, which is 5. Since the planning horizon is 7 days, the discharge time window is between 5 and 7 days. We assume that we have one overnight resource and no day resource. The overnight resource is required by both patients. Therefore, only one patient can be admitted and thus, only one admission decision variable is equal to 1. Table 2 in the Technical Appendix gives the algebraic solution of the bed allocation example.

Results and Discussion

In the following, a computational and economic analysis of the scheduling problem is provided.

Data and instance generation

We evaluated our model using data from a mid-sized hospital in the vicinity of Munich, Germany. We joined the hospital data with data from the German institute for reimbursement in hospitals [8]. The latter contains information on the DRG attributes such as low and high length of stay (LOS) trim points, fixed revenue as well as per day reduction and addition. We learned the classifier from January 2011 until June 2011. We ran the simulation experiments from July 1, 2011 until the rest of the year, where each day is solved independently.

Classification results

The classification results of our experiments reveal that the DRG grouper could not classify any of the patients' DRG before admission. One explanation for this phenomenon is that the grouper cannot deal with free-text information. A more detailed analysis revealed that the "error-DRG" 960Z was always assigned to the patients. In contrast, the true positive rate of the Naïve Bayes approach was, on average, 25.3%.

Contribution margin analysis

Our contribution margin analysis revealed that scheduling patients using their DRG, as classified by Naïve Bayes, can substantially increase the hospital's contribution margin, depending on the input parameters of the model and their uncertainty. Our results revealed up to 17% improvement in contribution margin, depending on the planning day. A more detailed analysis reveals that patients classified by the DRG grouper may not be admitted at all because they compete for scarce resources with those patients whose contribution margin is obtained using the Naïve Bayes-classified DRG.

Discussion

The connection between our early DRG classification and the patient admission and scheduling problem should be seen as preliminary results of a sensitivity analysis regarding the extent to which better DRG prediction leads to more effective resource allocation. These models and methods can now be embedded in a real-time decision support system that can inform a decision maker in a hospital about the value of early availability of information about patients to increase classification accuracy and, as a consequence, lead to better resource allocation decisions. This approach for early classification of patients' DRG is not only applicable for elective patients but can also be used for emergency patients, as shown in [6].

Another result of our decision support system based on the mathematical model is that a decision maker can be better informed about patients who would be admitted or scheduled based on a hospital's patient admission and scheduling policy. Furthermore, the decision variables of the model could also reveal that admitting and scheduling patients based on the model's recommendations may lead to higher contribution margins. Identifying these opportunities can also be supported by the approach and tool discussed in this paper which can be included in the platform of Gartner and Padman [7].

Finally, these models and methods provide a rigorous, systematic and evidence-based approach to integrating and analyzing financial objectives of healthcare delivery organizations with operational decisions that are driven by clinical requirements as well as planning and resource allocation constraints.

Conclusion

In this paper we have presented a novel approach to link classification methods with a discrete optimization model for the problem of scheduling elective patients hospital-wide. The objective is to maximize the contribution margin for elective patients using real-world data from a mid-sized hospital. The results show that if hospitals pursue an objective of making admission decisions based on DRGs, basic machine learning techniques such as Naïve Bayes can lead to more efficient resource allocation decisions as compared to the use of a hospital's current approach using a DRG grouper.

Several streams deserve further research. In order to make the model more applicable in hospitals, our goal is to incorporate overtime flexibility into the model and to consider more specific assignments of patients to a variety of resources, for example, human resources such as surgical teams. Following the approach of Gartner and Arnolds [3], we will also incorporate uncertain clinical pathways into the scheduling process. Preliminary results on a simplified model are shown in prior work [4]. Furthermore, we will evaluate other machine learning techniques such as decision trees, Bayesian networks and logistic regression in combination with the early DRG classification and resource allocation task. Another task is to break down the results into a length of stay analysis, and an evaluation of the confusion matrix in combination with admission and scheduling decisions.

Technical Appendix

Sets, indices, parameters and decision variables

Table 1 provides an overview of the sets, indices, parameters and decision variables.

Table 1 – Sets, indices, parameters and decision variables

Set	Description
A	Activities, Attributes
A_p	Activities for patient $p \in P$
D	DRGs
E_p	Precedence relations
P	Patients
R	Resources
R^d	Day resources
R^n	Overnight resources
T	Days
W_i	Time window for activity $i \in A$
Index/parameter	**Description**
α_p	Admission date of patient $p \in P$
$a \in A$	Attribute
b_p	Specialty requirement of patient $p \in P$
d	DRG
d_i^*	DRG to which instance i is assigned
$d_{i,j}^{\min} \geq 0$	Minimum time lag corresponding to precedence relation $(i,j) \in E_p$
E_i	Earliest day to schedule activity $i \in A$
i	Instance
$i \in A$	Activity
(i,j)	Precedence relation between activity $i \in A$ and $j \in A : i \neq j$
L_i	Latest day to schedule activity $i \in A$
$p \in P$	Patient
$p(d)$	Prior probability of DRG d
$p(v_{i,a} \mid d)$	Conditional probability of DRG d given instance i's value of attribute a
$\pi_{p,t}$	Contribution margin for patient $p \in P$ discharged at day $t \in W_{\phi_p}$
ϕ_p	Discharge activity for patient $p \in P$
$r_{i,k}$	Resource requirement of activity i from resource k
$t \in T$	Day
$v_{i,a}$	Value of instance i's attribute a
Decision variable	**Description**
$x_{i,t}$	1 if activity i is scheduled for day t, 0 otherwise
z_p	1 if patient p is admitted, 0 otherwise

Mathematical program

$$Maximize \sum_{p \in P} \sum_{t \in W_{\phi_p}} \pi_{p,t-\alpha_p} \cdot x_{\phi_p,t} \tag{2}$$

Subject to

$$\sum_{t \in W_j} t \cdot x_{j,t} \geq \sum_{t \in W_i} t \cdot x_{i,t} + d_{i,j}^{\min} \cdot z_p \qquad \forall p \in P, (i,j) \in E_p \tag{3}$$

$$\sum_{t \in A : t \in W_i} r_{i,k} \cdot x_{i,t} \leq R_{k,t} \qquad \forall k \in R^d, t \in T \tag{4}$$

$$\sum_{p \in P : b_p = k, t \geq \alpha_p} \left(z_p - \sum_{\tau = E_{\phi_p}}^{\min\{t, L_{\phi_p}\}} x_{\phi_p,\tau} \right) \leq R_{k,t} \qquad \forall k \in R^n, t \in T \tag{5}$$

$$\sum_{t \in W_i} x_{i,t} \leq 1 \qquad \forall i \in A \tag{6}$$

$$\sum_{i \in A_p, t \in W_i} x_{i,t} = |A_p| \cdot z_p \qquad \forall p \in P \tag{7}$$

$$x_{i,t} \in \{0,1\} \qquad \forall i \in A, t \in W_i \tag{8}$$

$$z_p \in \{0,1\} \qquad \forall p \in P \tag{9}$$

A bed allocation example

An example with two patients is presented in Table 1.

Table 2 – A bed allocation example for two patients

t	1	2	3	4	5	6	7
$x_{\phi_1,t}$	-	-	-	-	0	1	0
$\sum_{b_1=1, t \geq 1} \left(z_1 - \sum_{\tau = E_{\phi_1}}^{\min\{t, L_{\phi_1}\}} x_{\phi_1,\tau} \right)$	1	1	1	1	1	1	1
$x_{\phi_2,t}$	-	-	-	-	0	0	0

$\sum_{b_2=1, t \geq 1} \left(z_1 - \sum_{\tau=E_{\phi_2}}^{\min\{t, L_{\phi_2}\}} x_{\phi_2,\tau} \right)$	0	0	0	0	0	0	0

In this example, we assume that we have a planning horizon of $T = \{1,2,...,7\}$ days and that both patients $p = 1$ and $p = 2$ have an admission date $\alpha_1 = \alpha_2 = 1$ and a discharge time window of $W_{\phi_1} = W_{\phi_2} = \{5,6,7\}$. Moreover, we assume that we have one overnight resource $R^n = \{1\}$ and, for simplicity, no day resource $R^d = \{\ \}$. The overnight resource is required by both patients. Accordingly, the bed requirements for both patients come up to $b_1 = b_2 = 1$. Assume, that patient $p = 2$ cannot be admitted and therefore decision variable $z_2 = 0$. The table reveals that a bed is allocated for patient $p = 1$ from his admission in period $t = 1$ until his discharge in period $t = 6$. In contrast, patient $p = 2$ does not allocate a bed since he is not admitted and, because of constraints (7) he is not discharged.

This study is exempt from IRB because it analyzes de-identified data.

Acknowledgements

The authors sincerely thank the anonymous referees for their careful review and excellent suggestions for improvement of this paper.

References

[1] Organisation of economic co-operations and development, http://stats.oecd.org/Index.aspx?DataSetCode=SHA (last accessed: December 20th, 2014).

[2] Busse R, Geissler A, Quentin W, and Wiley M. Diagnosis-related groups in Europe. McGraw-Hill, Berkshire, 2011.

[3] Gartner D and Arnolds IV. Learning clinical pathways for downstream patient flow planning. Proceedings of 2014 INFORMS Workshop on Data Mining and Analytics (DMA 2014), 2014.

[4] Gartner D, Arnolds I and Nickel S. Improving hospital-wide patient scheduling decisions by clinical pathway mining. MEDINFO 2015 – 15th World Congress on Health and Biomedical Informatics, 2015.

[5] Gartner D and Kolisch R. Scheduling the hospital-wide flow of elective patients. Eur J Oper Res 2014: 223(3): 689-699.

[6] Gartner D, Kolisch R, Neill DB, and Padman R. Machine learning approaches for Early DRG Classification and Resource Allocation. INFORMS J on Computing: Accepted for publication, 2015.

[7] Gartner D and Padman R. E-HOSPITAL – An Environment for Hospital Operations and Services Planning using Information Technology and Algebraic Languages. MEDINFO 2015 – 15th World Congress on Health and Biomedical Informatics, 2015.

[8] Schreyögg J, Stargardt T, Tiemann O, and Busse R. Methods to determine reimbursement rates for diagnosis related groups (DRG): A comparison of nine European countries. Health Care Manag Sci 2006: 9(3): 215-223.

[9] Schreyögg J, Tiemann O, and Busse R. Cost accounting to determine prices: How well do prices reflect costs in the German DRG-system? Health Care Manag Sci 2006: 9(3) 269-279.

[10]Stabile M, Thomson S, Allin S, Boyle S, Busse R, Chevreul K, Marchildon G, and Mossialos E. Health care cost containment strategies used in four other high-income countries hold lessons for the United States. Health Affairs 2013: 32(4): 643-652.

[11] Anderson GF, Willink A, and Osborn R, Reevaluating "Made in America" — Two Cost-Containment Ideas from Abroad. N Engl J Med, 2013: 368(24): 2247-2249.

[12] van de Klundert J, Gorissen P, and Zeemering S. Measuring clinical pathway adherence. J Biomed Inf, 2010: 43(6): 861-872.

Address for correspondence

Daniel Gartner
The H. John Heinz III College
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
dgartner@andrew.cmu.edu