A data mining approach for estimating patient demand for health services

Ionut ȚĂRANU
Bucharest University of Economic Studies
E-mail: ionut.taranu@gmail.com

Abstract: *The ability to better forecast demand for health services is a critical element to maintaining a stable quality of care. Knowing how certain events can impact requirements, health-care service supplier can better assign available resources to more effectively treat patients' needs.*
The embodiment of data mining analytics can support available data to identify cyclical patterns through relevant variables, and these patterns provide actionable information to adequate decision markers at health-care structures.
The request for health-care services can be subject to change from time of year (seasonality) and economic factors. This paper exemplifies the efficacy of data mining analytics in identifying seasonality and economic factors as measured by time that affect patient demand for health-care services. It incorporates a neural network analytic method that is applied over a readily available dataset. The results indicate that day of week, month of year, and a yearly trend significantly impact the demand for patient services.

Keywords: *Data mining, nueronal networks, decision support systems, health care IT .*

1 Introduction

There is increasing acknowledgment of the worth of information that exists in data resources in institution through industry sectors. Tendency, relationships between variables, and repeated models all may exist in data bases and provide accurate and deep descriptions of different processes and increase the capacity to predict and generate quantitative patterns that facilitate decision support for specialists and practitioners. A vital element to identifying patterns and relationships and generating models that facilitate simulations are multivariate techniques. A prominent field in the multivariate arena imply data mining methods that include mathematical functions and algorithms that process data resources in order to extract actionable facts for decision makers. This process of information extraction through data mining is often referred to as knowledge discovery [1] or, in other words, the recognition of valuable information that enhances facts, information, and skills for those who make decisions. The notion of leveraging data resources with mining methods to augment decision making via know-how discovery is becoming an acute component of organizational capability bearing in mind the evolving era of big and new data resources. Data resources are increasing every year in the light of the introduction of new technologies across industry sectors.

The health-care industry is confronting a meaningful growth in data resources due to the continuous progress of the digital age. The creation of electronic medical records, the process of e-prescribing, medical devices that automatically download patient elements, and the increased usage of information systems at the private practitioner and hospital and health systems level are easing the initiation of vast resources that can supply information to increase yield in a number of applications. Data mining analytics are being applied in the health-care industry across a mixture of areas. Some of these include the analysis of workflow operations of large health-care provider companies that include studies that examine the drivers of patient duration of stay, patient request and bottlenecks in emergency room throughput, and patient satisfaction rates. Other sectors involve risk bedding applications or the better identification of patient populations at danger of developing chronic disease. Ultimately, semantic mining applications are being applied to electronic health records to better comprehend treatment and outcomes and patient diagnosis. One particular data mining technique involves the use of neural networks, which is an approach that include algorithms that processes historical data to detect both linear and non-linear patterns. The resulting models can then be used to lead 'what if' simulations on out of sample data, new data, and prognosis data. Neural networks have been used in a variety of industry applications that include the prediction of bank failures, traffic patterns, and even precipitations [2][3]. Neural networks and multivariate techniques are also being incorporated in the health-care industry to aid in knowledge discovery in a number of applications that comprise treatment efficaciousness [4] and general operations of health-care agencies [5] and patient throughput in emergency rooms.

2. Advanced analytics and health-care services

A significant factor in obtaining increased performance in providing services in any industry is the capability to better perceive what conduct demand for these services. With this information, decision makers can more specifically enforce the best or most favorable number of resources that are necessary to satisfy different quantity of request. Predictive analytic methods can improve the accuracy of estimating patient demand for organizations that enhance health-care services. Through analyzing patient services data in the readily available database, it was determined that seasonality factors together with general macroeconomic course had noteworthy effects on the demand for health-care services. With this information, decision makers can more precisely apply existing resources to satisfy patient expectations, and better handle costs while providing a more consistent service. In the quest to increase efficiencies, companies detect a process or functional zone that can be improved with respect to resource allocations that execute some type of task.
Decision makers study the situation at hand and regulate employee pools, technological infrastructure, and compatible operations.

2.1. Data mining and predictive modeling in patient centered decision support

Health-care informatics techniques are crucial in awareness and supporting health-care delivery components. Data mining and predictive modeling techniques are fundamental to this because of significant improvements in information technology as well as data collection and aggregation of disparate data sources.
Medical diagnostic decision support (MDDS) systems have been used for decades. These systems have been developed because it is well-known that health-care providers are frequently asked to make crucial clinical assessment based on imprecise and/or deficient patient information. Inadequate information leads to faults, which can dramatically influence quality of care. For example, treatment for patients suffering from diabetes is exacerbated by the presence of comorbid conditions, social support challenges, and poor medication adherence. . This type of electronic decision support system ensures the appropriate implementation of evidence-based chronic care models.
Data mining methods have been used to identify the socio-demographic, physical, and psychological factors most important to the early detection and treatment of serious health-care conditions. Penny & Smith [6] explored data mining techniques to improve the quality of life of patients suffering from irritable bowel syndrome (IBS). This longitudinal cohort study examined logistic regression,

classification, and neural network models. These models demonstrated that IBS severity, psychological morbidity, marital status, and employment status significantly influenced a patient's health-related quality of life. These results provide the best information to afford better assessment and management of patients with IBS.

Other studies have examined data mining methods to improve the accuracy of diagnostic systems based on information derived from multiple, disparate data sources as well as recognition of the uniqueness of health-care data mining methods and techniques [7]. In addition, with advances in information technology, it is now possible to combine data from electronic medical records with human knowledge (i.e., expert information) to optimize the accuracy of diagnostic systems. Prediction of the onset of liver cancer [8], classification of malignant colorectal tumors and abnormal livers [9], and prediction of mortality of patients with cardiovascular disease [10] are now commonplace.

Other important applications of health-care data mining and predictive modeling techniques are resource allocation and request management in the emergency department and hospital setting. Sun et al [11] developed forecasting models to determine the probability of a hospital admission based on information collected at the point of emergency department triage. Examining 2 years of hospital data collected by nurses from emergency department patients at the point of triage, regression models were developed to determine the strongest factors in precisely predicting a patient's immediate inpatient admission from the emergency department. Outside of the obvious admission criteria (e.g., heart attack, life-threatening trauma), it is not always clear that a patient will be admitted at the point of emergency department triage for conditions such as respiratory infections, pleurisy, or orthopedic concerns. The results from this study demonstrated that age, patient acuity category, and emergency department arrival mode were the strongest predictors for hospitalization. These predictive models, if used at the point of triage, could be used for early admission planning and resource challenges faced by inpatient and acute care facilities. Similarly, there have been several studies that demonstrate the effectiveness of data mining techniques in forecasting hospital admissions, returns to the emergency department, demand for specific illnesses, same-day admissions, and emergency department demand [12],[13],[14],[15].

2.2 Seasonality and estimating the demand

The notion of identifying repetitive or cyclical direction in time for demand of particular processes is frequently referred to as identifying seasonal patterns of demand.

Traditionally, companies apply analytics to establish two main sources of information concerning seasonality: whether demand for their products or services has seasonal patterns (e.g. do their sales increase or decrease according to a particular point in time on a repetitive basis) and, if seasonality exists, what is the size of the change in demand according to a particular point in time. Prior research has shown that seasonality effects for patient demand for healthcare exist. For example, the 'winter effect' has been cited as being associated with increases in depression-related ailments [15],[16]. Studies have also concluded that other factors such as general economic distress (e.g. unemployment, financial stress) drive demand for mental health services [18],[19].

Analytic techniques have been used to model the impact of seasonality on patient request for health services in order to better predict future demand and allocate resources consequently. Existing research incorporating calendar-based data has concluded that seasonality provides valuable decision support for the estimation of patient demand for urgent care clinics and emergency room facilities. Step-wise linear regression was applied to daily patient volume, which was matched with calendar data (e.g. day of week and month of year) and weather data to forecast the number of patients searching urgent care [5]. The results indicated that regression models incorporating calendar data were useful in estimating future patient demand while weather data only provided peripheral improvement to the analysis. Time series methods, linear regression, and neural network methods incorporating daily patient visits and calendar data have been utilized to study patient visits to emergency room facilities [20],[21]. This application is seen as particularly useful in helping alleviate overcrowding and enhance staffing and patient throughput by providing predictive information of patient demand [22].

2.3. Data and analytic methodology

In this chapter, we summarize the data collected from outpatient (patient who receives medical treatment without being admitted to a hospital) clinics that included ENT, dermatology, paediatrics, orthopaedics, and OBGYN clinics.

Table 1 classify the monthly and intra-week seasonality indices. These indices are calculated by the ratio of recurrent demand to mean demand. The coefficient of variation (Cv= σ / μ) is calculated as a relative measure on the levels of seasonality through the clinics. As one would expect, differences can be found in both the patterns and the levels of monthly seasonality for various types of specialties. For example, request is increased during the summer time for dermatology, whilst the opposite is true for ENT. Then again, from the the intra-week models presented in Table 2 we can see that are less likely to be affected by the type of specialty. A general pattern reveals peaks on Mondays, followed by Thursdays next. We can usually notice lower demands Tuesdays and Wednesdays. Intra-day variations existed with peaks around mid-mornings and mid-afternoons.

Table 2.1 .Seasonality data - Monthly seasonal indices

Monthly seasonal indices	ENT	Orthopaedics	Paediatrics	Dermatology	OBGYN
January	1.557	1.326	1.241	0.919	1.250
February	1.413	1.285	1.089	1.027	1.293
March	1.654	1.123	1.093	1.459	1.116
April	1.307	1.164	1.048	1.363	1.178
May	0.953	1.367	1.108	1.303	1.256
June	0.999	1.184	1.205	1.952	1.122
July	0.840	1.245	1.044	1.134	0.916
August	0.908	1.529	1.037	1.183	1.328
September	1.052	1.266	1.395	1.496	1.122
October	1.307	1.225	1.486	1.063	1.384
November	1.365	1.034	1.422	0.866	1.428
December	1.403	1.014	1.593	0.998	1.367
Coefficient of variation (Cv)	0.269	0.143	0.196	0.306	0.145

Table 2.2 Seasonality data - Intra-week seasonal indices

Intra-week seasonal indices	ENT	Orthopaedics	Paediatrics	Dermatology	OBGYN
Monday	1.498	1.400	1.455	1.431	1.380
Tuesday	1.130	1.239	1.054	1.270	1.220
Wednesday	1.170	1.062	1.110	0.891	0.988
Thursday	1.225	1.215	1.211	1.446	1.281
Friday	1.124	1.231	1.319	1.076	1.280
Coefficient of variation (Cv)	0.126	0.119	0.161	0.226	0.146

Neural network analysis and results

The neural network methodology in this case allude to the utilization of complex computer algorithms that detect existing models and relationships within historical data. The neural network framework used in this analysis incorporates a multilayered perceptron[23]with a feedforward backpropagation testing function [24] .

The neural network modeling process begins with an input layer that includes nodes that correspond to each independent (driver) variable.

Driver variables are assigned weights by the algorithm, where the weighted sum of these inputs is passed into a squashing function in the hidden layer where non-linear calculations are performed on the variables relative to the dependent variable. The combined results in the input and hidden layers are passed to an output layer and compared with the historical dependent variable. Weights for variables are estimated by the backpropagation training method.

The final model is a set of code that involves a allowance or adjustment made in order to take account of special circumstances or compensate for a distorting factor scheme for independent/driver variables. Neural networks can be compared with regression analysis, with a major differentiator being that the n-net approach is based in algorithmic processing that incorporates a dynamic weighting mechanism.

3. Conclusions

These advanced analytic methods go across simple recognition of retrospective capabilities of prime reporting and supply decision markers with quantitative models that describe relationships among variables underpinning processes. These models provide simulation capabilities to project eventual outcomes given customization to process sariable inputs.

More unpretentiously put, analytic methods such as neural networks process historical data and determine whether there are reliable steadiness in the the rate at which things occurs or is repeated over a particular period of time and magnitude of occurrences in that data [25]. In this case do Mondays or Tuesdays of every week or particular months over the 3 years entail significant trends/patterns regarding demand for patient services.

The yield for the multivariate approach could return practical value information for hospital staffing office. The output could spot whether a particular day of the week consistently experiences +/- average demand, and would also provide an assessment of the detailed level of the demand. With this knowledge, health-care staffing operations can better keep suitable clinicians on a daily basis with greater precision to facilitate solid care for patients. In the case at hand, significant seasonal patterns were identified.

References

[1] Fayyad U, Shapiro G and Smyth P (1996) From data mining to knowledge discovery in databases. Artificial Intelligence 13(3), 37–54.

[2] Tam K and Kiang M (1992) Managerial applications of neural networks: the case of bank failure predictions. Management Science 38(7), 926–947. | Article |

[3] French M, Krajewski R and Cuykendall R (1992) Rainfall forecasting in space and time using a neural network. Journal of Hydrology 137(1–4), 1–31. | Article

[4] Lisboa P et al (2008) Time-to-event analysis with artificial neural networks: an integrated analytical and rule-based study for breast cancer. Neural Networks 21(2–3), 414–426. | Article | PubMed |

[5] Batal H, Tench J, McMillan S, Adams J and Mehler PS (2001) Predicting patient visits to an urgent care clinic using calendar variables. Academic Emergency Medicine 8(1), 48–53. | Article | PubMed |

[6] Penny KI and Smith GD (2012) The use of data-mining to identify indicators of health-related quality of life in patients with irritable bowel syndrome. Journal of Clinical Nursing 21: 2761–2771. | Article | PubMed |
[7] Cios KJ and More GW (2002) Uniqueness of medical data mining. Artificial Intelligence in Medicine 26(1–2), 1–24. | Article | PubMed |
[8] Kuo MH, Hung CM, Barnett J and Pinheiro F (2012) Assessing the feasibility of data mining techniques for early liver cancer detection. Studies in Health Technology and Informatics 180: 584–588. | PubMed
[9] Gao P et al (2012) Which is a more accurate predictor in colorectal survival analysis? Nine data mining algorithms vs. the TNM staging system. PloS One 7(7), e42015. Epub 25 July. | Article | PubMed |
[10] Austin PC, Lee DS, Steyerberg EW and Tu JV (2012) Regression trees for predicting mortality in patients with cardiovascular disease: what improvement is achieved by using ensemble-based methods? Biometrical Journal 54: 657–673. | Article | PubMed |
[11] Sun Y, Heng BH, Tay SY and Seow E (2011) Predicting hospital admissions at emergency department triage using routine administrative data. Academic Emergency Medicine 18(8), 844–850. | Article | PubMed |
[12] Peck JS, Benneyan JC, Nightingale DJ and Gaehde SA (2012) Predicting emergency department inpatient admissions to improve same-day patient flow. Academic Emergency Medicine 19(9), E1045–E1054. | Article | PubMed |
[13] LaMantia MA et al (2010) Predicting hospital admission and returns to the emergency department for elderly patients. Academic Emergency Medicine 17(3), 252–259. | Article | PubMed |
[14] Kudyba S (2012) Utilizing multivariate analytics for decision support to enhance patient demand forecasting. International Institute for Analytics October, 1–7.
[15] Jones SS et al (2009) A multivariate time series approach to modeling and forecasting demand in the emergency department. Journal of Biomedical Informatics 42(1), 123–139. | Article | PubMed |
[16] Lurie SJ, Gawinski B, Pierce D and Rousseau SJ (2006) Seasonal affective disorder. American Family Physician 74(9), 1521–1524. | PubMed |
[17] Fullerton KJ and Crawford V (1999) The winter bed crisis – quantifying seasonal affects on hospital bed usage. QJM 92(4), 199–206. | Article | PubMed |
[18] Catalano R (1991) The health effects of economic insecurity. American Journal of Public Health 18(9), 1448–1152.
[19] Dooley D, Prause J and Ham-Rowbottom K (2000) Underemployment and depression: longitudinal relationships. Journal of Health and Social Behavior 41(4), 421–436. | Article | PubMed |
[20] Holleman DR, Bowling RL and Gathy C (1996) Predicting daily visits to a walk-in clinic and emergency department using calendar and weather data. Journal of General Internal Medicine 15(4), 237–239. | Article |
[21] Jones SS, Thomas A, Evans RS, Welch SJ, Haug PJ and Snow GL (2008) Forecasting daily patient volumes in the emergency department. Academic Emergency Medicine 15(2), 159–170. | Article | PubMed |
[22] Wargon M, Guidet B, Hoang TD and Hejblum G (2009) A systematic review of models for forecasting the number of emergency department visits. Emergency Medicine Journal 26(6), 395–399. | Article | PubMed |

[23] Rumelhart D, Hinton G and Williams RJ (1986) Learning internal representations by error propagation. In Parallel Distributed Processing: Explorations in the Microstructure of Cognition. (Rumelhart DE, McClelland JL and Corporate PDP Research Group, Eds) MIT Press.
[24] Hinton G (1992) How Neural Networks Learn from Experience. Scientific American 267(3), 144–151. | Article | PubMed |
[25] Chu C and Zhang G (2003) A comparative study of linear and nonlinear models for aggregate retail sales forecasting. International Journal of Production Economics 86(3), 217–231. | Article |

Ionuţ Ţăranu graduated from the Faculty of Cybernetics, Statistics and Economic Informatics of the Academy of Economic Studies in 1996, having its Master degree on "Database support for business". At present is in the process of getting his title of doctor in economy in the specialty of "Soft-computing methods for early medical diagnosis". He has been an Assistant Professor for 4 years at "Titu Maiorescu" University and also for 4 years at Academy of Economic Studies from Bucharest. He published a series of articles, from which the most important are Applying ABCD Rule of Dermatoscopy using cognitive systems and ABCDE Rule in Dermoscopy – Registration and determining the impact of parameter E for evolution in diagnosing skin cancer using soft computing alghorithms.
Mr. Ţăranu is currently the General Manager of Stima Soft company. He has more than 15 years of experience as a project manager and a business analyst with over 13 years of expertise in Software development, Business Process Management, Enterprise Architecture design and Outsourcing services. He is also involved in research projects, from which the most relevant are:

- Development of an Intelligent System for predicting, analyzing and monitoring performance indicators of technological and business processes in renewable energy area;
- Development of an eHealth platform for improving quality of life and the personalization of therapy at patients with diabetes;
- Development of an Educational Portal and integrated electronic system of education at the University of Medicine and Pharmacy "Carol Davila" to develop medical performance in dermatological oncology field;