



Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 170 (2020) 1029–1034

Procedia
Computer Science

www.elsevier.com/locate/procedia

International Workshop on Hospital 4.0 (Hospital)
April 6-9, 2020, Warsaw, Poland

Predictive and Prescriptive Analytics in Healthcare: A Survey

João Lopes*, Tiago Guimarães, Manuel Filipe Santos

University of Minho, Centro Algoritmi, Braga, Portugal

Abstract

Over the years, health area has received numerous studies on how to improve its management and administration activities and, fundamentally, the Healthcare provided to its patients. Currently, there is an exponential growth of data in the health system. In this sense, it is crucial the implementation of technologies capable of using it in a beneficial way for the organization, helping it to fulfill its strategic objectives. Subsequently, this same data, if used correctly, has the capacity to assist the organization at an administrative level, as well as at the level of patient care, using predictive and optimization models capable of revolutionizing the current health system. Thus, this article aims to identify the advances that have been made in this area, focusing on the development of predictive and optimization techniques, applied in Health, and how these can improve the lives of managers, doctors and patients.

© 2020 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Peer-review under responsibility of the Conference Program Chairs.

Keywords: Healthcare; predictive; prescriptive; analytics; data mining; machine learning;

1. Introduction

Nowadays, the provision of hospital services depends on the efficient execution of processes, which includes a set of clinical or non-clinical activities, as well as different resources such as doctors or nurses, among others. They make the processes highly dynamic and increasingly complex. A system capable of improving, evaluating and preventing certain situations becomes an increasingly important point in health organizations, due to the objectives that each one has to fulfill, but mainly due to the care provided to patients. The growing availability of data and the quick development of these methods make possible to discuss how clinical and administrative decisions can be made easier,

* Corresponding author. Tel.: +351934373667.

E-mail address: lopesit@outlook.pt

in order to replace human intuition in certain areas, revealing information hidden behind all the data existing in the same organization [7].

These days, based on the revised literature, we can identify predictive analytics mostly as clinical decision support, with many studies involving risk and morbidity prediction of chest pain, heart attack, and other diseases. In contrast, the use of prescriptive analytics is found in public Healthcare, administration, and mental health [2]. In fact, hospital services generate various types of data, like structured data, for example from laboratory or radiology results, past medical history, medications lists, or allergy information, among other, and unstructured data, such as notes on treatment progress or other Healthcare provider reports. In addition, while medical data grows fast, computing power and technical ability can offer a remake of medical field in advanced ways that never were thought before. Thus, the integration of predictive and prescriptive analytics in Healthcare is something widely important, not only because of the capacity to support the data that are generated on a daily basis, but also because of the expectation of inserting the evolution of machine learning techniques in the improvement of the care provided to the community, without discarding the capacity to assist administrators in the management of strategic decisions and the positive contribution that can have in the functioning of the same organization [3] [8].

2. Predictive Analytics in Healthcare

2.1. Basic Concepts

Predictive analytics depends on machine learning, building algorithms in order to extract knowledge through existing data, combining them and, as a result, determine the future. These same algorithms can be divided into two main categories: Unsupervised Learning and Supervised Learning, as shown in Fig. 1.

Unsupervised Learning is mostly known for characteristic extraction, it means that they can be very useful in different types of analysis because they will automatically identify structure in data, while Supervised Learning is typically done in the context of classification, when we want to map input to output labels, or regression, when we want to map input to a continuous output. More recently, Semi-supervised Learning has been introduced as a hybrid between Unsupervised and Supervised Learning, and the algorithm is trained upon a combination of labeled and unlabeled data and is suitable for scenarios where there is a lack of results in certain contexts [7].

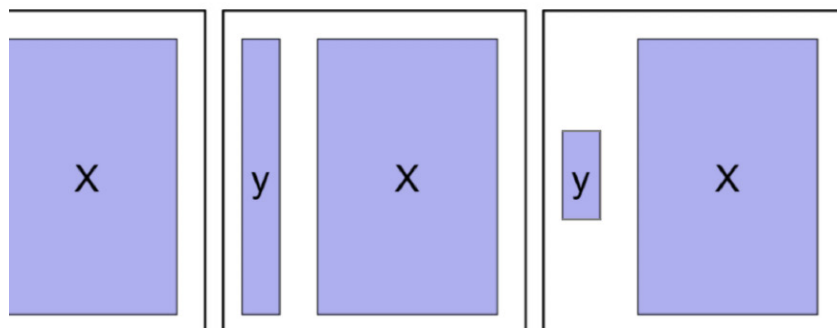


Figure 1 - Three Types of Machine Learning Algorithms. Withdrawn from [7]

The major Unsupervised Learning method is Clustering, a process of grouping similar entities. The goal is to find similarities in the data point and group similar data points together. Supervised Learning provides more clinically relevant results and, as one can see in Fig. 2, contain the most relevant techniques such as linear regression, logistic regression, naive bayes, decision tree, nearest neighbour, random forest, support vector machine and neural network.

We can make a clearly distinction in this two major categories: Unsupervised Learning can be used as part of the pre-processing step to reduce dimensionality or identify subgroups, summarizing and explaining data features, making

a significantly contribute in the final results. Supervised learning will make a greater contribution to the efficiency, both of processing and of the final results that are intended to be obtained, according to the objectives and the type of existing data [7].

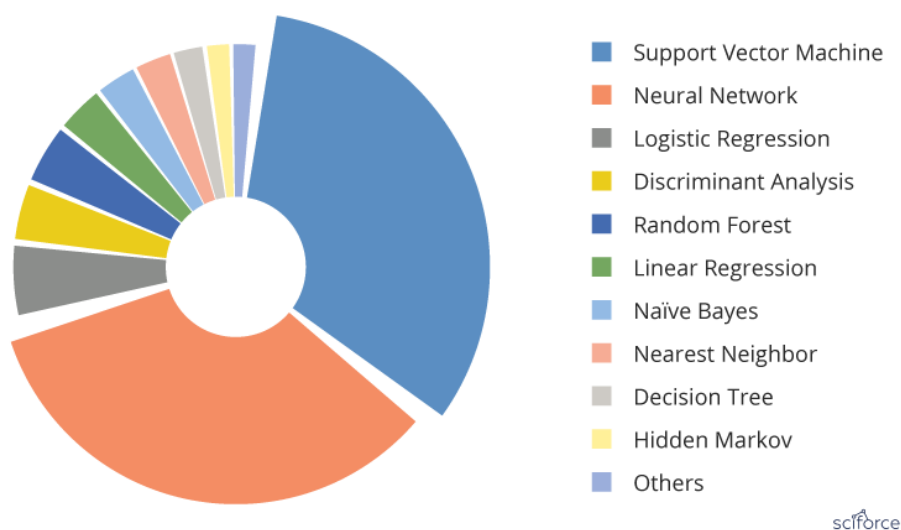


Figure 2 - Top Supervised Learning Machine Learning Methods. Withdrawn from [7]

2.2. Study Cases and Contributes

Some work has been developed in this area around predictive models, specially with clinical data, and how this studies can contribute positively in Healthcare.

With clinical data we can distinguish predictive research in cancer detection, in which many research projects want to detect cancer at an early stage, based on genetic and non-genetic factors. In one recent study, a hybrid system was developed with data mining algorithms, to build a predictive model capable of distinguish between breast cancer and non-breast cancer diagnoses and getting 100% accuracy in this model. In another study, conducted in Taiwan, a team of researchers utilized data mining techniques based on logistic regression to predict the survival rate among oral cancer patients, with 95.7% accuracy. In other perspective, another study use predictive models to distinguish three variables on patients with acute leukemia. Combining six algorithms, they can return disease stage, donor type and conditioning regimen that mostly contribute to the mortality risk of this patients. Exploring researches on heart diseases, there is a great potential to improve early detection accuracy, risk assessment, and ongoing symptom tracking through the careful application of predictive analytics. One recent study, made by University of California, classified patients as having or not having heart disease, using thirteen risk factor, with an accuracy of 93.02%. Another research team from Rice University used predictive analytics to identify early warning signs for a heart attack using scorecards [2].

A mention can be made to a study with the objective to predict drug side effects [3], with a development of a hybrid machine learning approach to predict drug side effects, categorize side effects into several intervals, adopt suitable strategies for each interval, and construct data models accordingly. These results showed a different perspective, achieve better predictive performances, analysing networks of associations between drugs and side effects. It also prove that some side effects must be carefully analysed and additional deterministic features need to be extracted.

These are just notable examples from the literature exemplifying that predictive analytics are present and highly marked in Healthcare and effectively will become an indispensable tool for clinicians, in interventions and improving patient's life.

3. Prescriptive Analytics in Healthcare

3.1. Basic Concepts

Prescriptive analytics is the next phase of predictive scenarios. In other words, this is the type of analytics that suggest action in different scenarios. It is highly associated to optimization problems and require sophisticated algorithms to find the optimum solution [10]. This type of problems can be solved with four distinct categories: Blind Search, Local Search, Search based on Population and Multi-objective Optimization.

The Blind Search assumes the exhaustion of all alternatives, ensuring that all solutions within the results space are tested and, subsequently, the certainty that the optimal solution is always found. In contrast to the Blind Search methods presented in the previous topic, the most modern optimization techniques are based on Local Search, from which new solutions are generated from existing solutions, including several methods that focus on a local neighborhood through a given initial solution. Population-based search tends to explore more distinct regions of the search space. As a consequence, more diversity can be achieved in terms of defining new solutions, which can be created not only by slightly changing each individual search point, but also by combining attributes related to two or more search points. Finally, the Multi-objective Optimization arises with the need to reconcile multiple objectives in a single algorithm, even more so if the achievement of some involves the loss of others, with the importance of something capable of achieving the greatest possible balance [2].

3.2. Study Cases and Contributes

After this context, a set of investigations are introduced in this area, applying the importance of mathematical and statistical models in recommending the best actions in certain contexts.

According to a research by University of Craiova [4], a real problem in Healthcare is how to allocate Healthcare resources because, without an efficient management, the Healthcare system will not be able to offer high quality patient care with affordable costs. One of the many issues is the appropriate use of beds. This study uses bed-occupancy data collected at the St. George's Hospital, in London, and reveal that a bed-occupancy exceeding 91% implicate a patient rejection rate around 1.1%. Besides, to determine a way to control this problem, encode a queue model based on genetic algorithms to optimize bed allocation and associated costs. The achievement of these goals provides a flexible and effective tool to simulate different possible scenarios, with recommended actions for each situation.

In another case study, according to Jagreet Kaur and Dr. Kulwinder Singh Mann [8], is suggest an architecture for enabling Artificial Intelligence based Healthcare analytics, applying prescriptive analytics techniques in this sector. To optimize the clinical process, they develop a mobile app with Artificial Intelligence, asking patients about their symptoms and provides information about their Health. This system uses natural Language processing and machine learning algorithms to create a map of patient condition and give a personalized experience. Moreover, the mobile app is prepared to use wireless sensors and transmit patient's vitals, using real-time alerts to alert nurses or doctors, respond more rapidly to unexpected changes.

One final appointment, based on a model for optimize neurology appointments in USA [1], suggest a resolution for patient no-shows and late cancellations, a common problem that cause financial and organization problems in Healthcare. In general, is created predictive models to assess risk of no-show, and then assign overbooking appointments utilizing those risks. Incorporating a final stage analytics, its propose a model, using predictive algorithms, such as Decision Tree, Random Forest or Naive Bayes, and for tuning the parameters of the learning algorithms and the balancing technique used grid search method along with cross-validation. This model can be very useful to the schedulers with high impact decisions, categorize high-risk patients to non-risk patients, planning alternatives for scenarios with certain behaviors can be taken into account.

4. Conclusion

Designing a hospital of the future, involves integrating both hospital services and various innovation technologies. The idea is, fundamentally, to create an intelligent and dynamic environment, with the aim of serving clinical and administrative interests, improving the quality of work of all those involved in a health entity with systems that can facilitate decision-making and, at the same time, contribute positively to the health and well-being of the patient.

According to what is reviewed in this article, and framing the focus areas of the article, it is possible to identify multiple cases of how an evolution can and should be done in Health. The integration of predictive Analysis can be extremely important in several areas of a health organization, from the clinical support that can be given in the discovery of how diseases spread as in the anticipation of an event so that better and more accurate Healthcare is given to the patient. At the administrative level, the entry of this area into the management of a Healthcare organization can have a significantly high impact, if we take into account that a better organization of the hospital can mean an increase in the provision of services to a growing community.

Prescriptive analysis becomes a complement to predictive analysis, being a final support for an entire intelligent system, which is sufficiently predictable to prevent scenarios and recommend the best actions. Thus, as it is possible to notice in the cases reported above, the introduction of mathematics and statistics in the development of algorithms makes possible to provide a hospital with a complete management of it. The optimization of existing resources may, for example, allow hospitals to shorten existing waiting lines, or also, with the integration of a system capable of identify the health problems prevent them from being solved outside the hospitals, so that there may be a greater availability to provide treatment to those who effectively need it.

For future work, it is planned the modeling of a possible architecture, which can integrate several areas and components related to this same research, with the objective of serving the clinical and administrative interests in Healthcare.

Acknowledgements

The work has been supported by FCT – Fundação para a Ciência e Tecnologia within the Project Scope: UID/CEC/00319/2020.

References

- [1] Ahmadi, Ehsan & Garcia-Arce, Andres & Masel, Dale & Reich, Eric & Puckey, Jason & Maff, Rebecca. (2019). A metaheuristic-based stacking model for predicting the risk of patient no-show and late cancellation for the neurology appointments. *IIE Transactions*.
- [2] Alharthi, Hana. (2018). Healthcare predictive analytics: An overview with a focus on Saudi Arabia. *Journal of Infection and Public Health*.
- [3] Ashfaq, Awais & Nowaczyk, Slawomir. (2019). Machine learning in healthcare -- a system's perspective.
- [4] Belciug, Smaranda & Gorunescu, Florin. (2016). A hybrid genetic algorithm-queueing multi-compartment model for optimizing inpatient bed occupancy and associated costs. *Artificial Intelligence in Medicine*.
- [5] Cortez, Paulo. (2014). *Modern Optimization with R, Use R! series*, Springer.
- [6] Islam, Md Saiful & Hasan, Md Mahmudul & Wang, Xiaoyi & Noor-E-Alam, Md & Germack, Hayley. (2018). A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining. *Healthcare*.
- [7] Jiang, Fei & Jiang, Yong & Zhi, Hui & Dong, Yi & Li, Hao & Ma, Sufeng & Wang, Yilong & Dong, Qiang & Shen, Haipeng & Wang, Yongjun. (2017). Artificial intelligence in healthcare: past, present and future. *BMJ*.
- [8] Kaur, Jagreet & Mann, Kulwinder. (2018). AI based HealthCare Platform for Real Time, Predictive and Prescriptive Analytics using Reactive Programming. *Journal of Physics: Conference Series*.
- [9] Lee, Wei-Po & Huang, Jih-Yuan & Chang, Hsuan-Hao & Lee, King-Teh & Lai, Chao-Ti. (2017). Predicting Drug Side Effects Using Data Analytics and the Integration of Multiple Data Sources. *IEEE Access*.
- [10] Mohammed, Emad & Far, Behrouz & Naugler, Christopher. (2014). Applications of the MapReduce programming framework to clinical big data analysis: Current landscape and future trends. *BioData mining*.

- [11] Rojas, Eric & Munoz-Gama, Jorge & Sepulveda, Marcos & Capurro, Daniel. (2016). Process Mining in Healthcare: A literature review. *Journal of biomedical informatics*.