KISDI
KOREA INFORMATION SOCIETY DEVELOPMENT INSTITUTE
Asian Journal of Information and Communications

Applications of big data and deep learning in healthcare industry from disease detection to cost reduction

M. Seifolahzade* Mazandaran university of science and technology
N. Chitgar** Mazandaran university of science and technology
R. Abbasnejad*** Mazandaran university of science and technology
M. Ahangaran**** Iran University of Science and Technology

Abstract

Big data analytics and deep learning are two high-focus of data science. A key benefit of deep learning is the analysis and learning of massive amounts of unsupervised data, making it a valuable tool for big data analytics where raw data is largely unlabeled and un-categorized. As the data keeps getting bigger, deep learning is coming to play a key role in providing big data predictive analytics solutions. Nowadays, many complex processes can generate big data, for example, there are a greater number of healthcare industry than ever before, collecting many terabytes of data per day. Today we use deep learning method for identifying metastatic breast cancer, recurrent neural cascade model for automated image annotation or predicting healthcare-associated infections. In this paper, we provide a brief overview of big data and deep learning, and some of its applications in the field of health care are mentioned.

Keywords: big data, deep learning, health care, cost reduction

1. Introduction

Deep learning and big data are two hottest trends in the rapidly growing digital world (Chen and Lin, 2014). The goal of machine learning is to enable a system to learn from the past or present and use that knowledge to make predictions or decisions regarding unknown future events (Landset et al., 2015). The general focus of machine learning is the representation of the input data and generalization of the learnt patterns for use on future unseen data. Deep learning algorithms are one promising avenue of research into the automated extraction of complex data representations (features) at high levels of abstraction (Najafabadi et al., 2015).

The meaning of the term "big data" is still the subject of some disagreement, but it generally refers to data that is too big or too complex to process on a single machine (Landset et al., 2015). Big data represents the general realm of problems and techniques used for application domains that collect and maintain massive volumes of raw data for domain-specific data analysis. Modern data-intensive technologies as well as increased computational and data storage resources have contributed heavily to the development of big data science. Mining and extracting meaningful patterns from massive input data for decision making, prediction, and other inference is at the core of big data analytics (Najafabadi et al., 2015).

* Corresponding author: mseif@ustmb.ac.ir
** nchitgar@ustmb.ac.ir
*** fanavarienovin@gmail.com
**** ahangaran@iust.ac.ir

According to the latest traffic forecast report by Cisco Systems (Chen and Lin, 2014), half a billion mobile devices were globally sold in 2015, and the mobile data traffic grew by 74% generating 3.7 exabytes (1 exabyte = 10^{18} bytes) of mobile data per month. Mobile big data (MBD) is a concept that describes a massive amount of mobile data which cannot be processed using a single machine. Deep learning is a solid tool in MBD analytics (Alsheikh et al., 2016).

Technology based companies such as Google, Yahoo, Microsoft, DoD and Amazon have collected and maintained data that is measured in exabyte proportions or larger. Moreover, social media organizations such as Facebook, YouTube, and Twitter have billions of users that constantly generate a very large quantity of data. Various organizations have invested in developing products using big data analytics to address their monitoring, experimentation, data analysis, simulations, and other knowledge and business needs, making it a central topic in data science research (Najafabadi et al., 2015).

For example, researchers need to use big data to discover new drugs. Marketers need to use social networks, mobile, geo-location, and sensor data to reach more customers. The United States National Security Agency (NSA) needs to process the exabytes of data collected over the internet in the Utah Data Center (Zhao, MacKinnon and Gallup, 2015).

2. Overview of big data

Big data is a term that describes the large volume of data – both structured and unstructured – that inundates a business on a day-to-day basis. But it's not the amount of data that's important. It's what organizations do with the data that matters. Big data can be analyzed for insights that lead to better decisions and strategic business moves.

Big data generally refers to data that exceeds the typical storage, processing, and computing capacity of conventional databases and data analysis techniques. As a resource, big data requires tools and methods that can be applied to analyze and extract patterns from large-scale data. The rise of big data has been caused by increased data storage capabilities, increased computational processing power, and availability of increased volumes of data, which give organization more data than they have computing resources and technologies to process. In addition to the obvious great volumes of data, big data is also associated with other specific complexities.

New methodologies and technologies are required for analyzing and handling big data. Developing efficient learning methodologies and techniques is crucial to analyze and understand big data so as to extract useful knowledge. Feasible and potential strategies for analyzing and processing big data may include (1) parallel and distributed computation, (2) instance selection and dimensionality reduction, and (3) incremental learning.

3. Why is big data?

The importance of big data doesn't revolve around how much data you have, but what you do with it. You can take data from any source and analyze it to find answers that enable 1) cost reductions, 2) time reductions, 3) new product development and optimized offerings, and 4) smart decision making. When you combine big data with high-powered analytics, you can accomplish business-related tasks such as:

. determining root causes of failures, issues and defects in near-real time.
. generating coupons at the point of sale based on the customer's buying habits.
. recalculating entire risk portfolios in minutes.

. detecting fraudulent behavior before it affects your organization.

Big data characteristics can be described by “6Vs”. They are: Volume, Velocity, Variety, Value, Variability and Veracity.

Volume: this means data size such as terabytes (TB: approximately 10^{12} bytes), petabytes (PB: approximately 10^{15} bytes) and zettabytes (ZB: approximately 10^{21} bytes), etc.

Velocity: data is generated at a high speed.
Variety: this represents all types of data such as structured data from relational tables, semi-structured data from key-value web clicks and unstructured data from email messages, articles and streamed video and audio, etc.
Value: it is defined by the added-value that the collected data can bring. It refers to the value that the data adds to creating knowledge. There is some valuable information somewhere within the data.
Variability: it refers to data changes during processing and lifecycle. Increasing variety and variability also increases the attractiveness of data and the potentiality in providing unexpected, hidden and valuable information.
Veracity: it includes two aspects: data consistency (or certainty) and data trustworthiness. Data can be in doubt: incompleteness, ambiguities, deception and uncertainty due to data inconsistency, etc.

Challenges: big data analytics faces a number of challenges, some key problem areas include: data quality and validation, data cleansing, feature engineering, high-dimensionality and data reduction, data re-presentations and distributed data sources, data sampling, scalability of algorithms, data visualization, parallel and distributed data processing, real-time analysis and decision making, crowdsourcing and semantic input for improved data analysis, tracing and analyzing data provenance, data discovery and integration, parallel and distributed computing, exploratory data analysis and interpretation, integrating heterogeneous data, and developing new models for massive data computation.

4. Overview of deep learning

Today, the problem of big data collections is often solved through distributed storage systems, which are designed to carefully control access and management in a fault-tolerant manner. One solution for the problem of big data objects in machine learning is through parallelization of algorithms.

Machine learning algorithms ‘learn’ to predict outputs based on previous examples of relationships between input data and outputs (called training data). A model of the relationship between inputs and outputs is gradually improved by testing its predictions and correcting when wrong. Machine learning is a set of computerized techniques for recognizing patterns in data. It is a way of generating something like the ‘line of best fit’. It’s useful to automate this process when the data has many features and is very complex.

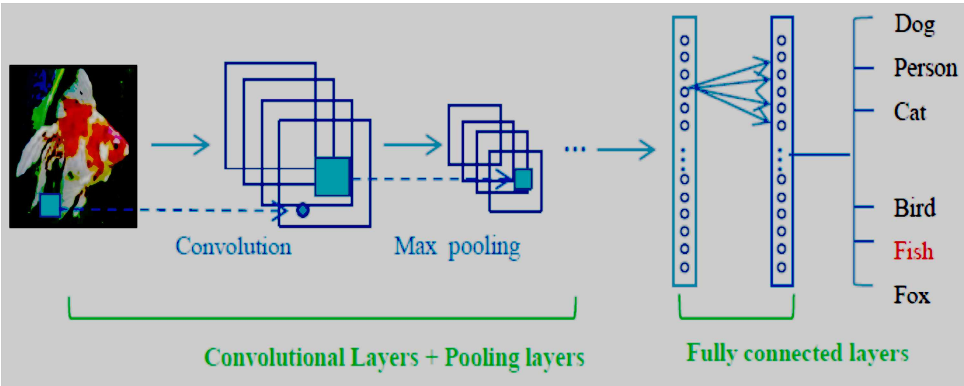
Deep learning refers to a set of machine learning techniques that learn multiple levels of representations in deep architectures. A key concept underlying deep learning methods is distributed representations of the data, in which a large number of possible configurations of the abstract features of the input data are feasible, allowing for a compact representation of each sample and leading to a richer generalization.

Deep learning algorithms are actually deep architectures of consecutive layers. Each layer applies a nonlinear transformation on its input and provides a representation in its output. The objective is to learn a complicated and abstract representation of the data in a hierarchical manner by passing the data through multiple transformation layers. The sensory data (for example pixels in an image) is fed to the first layer. Consequently the output of each layer is provided as input to its next layer. The two deep architectures are: convolutional neural networks (CNNs) and deep belief networks (DBNs).

5. Convolutional neural networks

A typical CNN is composed of many layers of hierarchy with some layers for feature representations (or feature maps) and others as a type of conventional neural networks for classification. It often starts with two altering types of layers called convolutional and subsampling layers: convolutional layers perform convolution operations with several filter maps of equal size, while subsampling layers reduce the sizes of proceeding layers by averaging pixels within a small neighborhood.

Figure 1. A sample of CNNs



6. Deep belief networks

Conventional neural networks are prone to get trapped in local optima of a non-convex objective function, which often leads to poor performance. Furthermore, they cannot take advantage of unlabeled data, which are often abundant and cheap to collect in big data. To alleviate these problems, a deep belief network (DBN) uses a deep architecture that is capable of learning feature representations from both the labeled and unlabeled data presented to it. It incorporates both unsupervised pre-training and supervised fine-tuning strategies to construct the models: unsupervised stages intend to learn data distributions without using label information and supervised stages perform local search for fine tuning.

7. Deep learning role in big data

As stated previously, deep learning algorithms extract meaningful abstract representations of the raw data through the use of a hierarchical multi-level learning approach, where in a higher-level more abstract and complex representations are learnt based on the less abstract concepts and representations in the lower level(s) of the learning hierarchy.

While deep learning can be applied to learn from labeled data if it is available in sufficiently large amounts, it is primarily attractive for learning from large amounts of unlabeled/unsupervised data, making it attractive for extracting meaningful representations and patterns from big data.

Other useful characteristics of the learnt abstract representations by deep learning include:

1. Relatively simple linear models can work effectively with the knowledge obtained from the more complex and more abstract data representations.
2. Increased automation of data representation extraction from unsupervised data enables its broad application to different data types, such as image, textural, audio, etc.
3. Relational and semantic knowledge can be obtained at the higher levels of abstraction and representation of the raw data.

8. Application of deep learning and big data in healthcare industry

Deep learning and big data have been successfully applied to many fields, such as image recognition, speech recognition, and machine translation, and embedded into industrial systems, like AlphaGo developed by Google Deep Mind. The success of deep learning has brought new insights into the medical domain where there are large quantities of data available.

Biomedical informatics is “the inter-disciplinary field that studies and pursues the effective uses of biomedical data, information, and knowledge for scientific inquiry, problem solving, and decision making, driven by efforts to improve human health”. Electronic Health Records (EHR) is a typical kind of biomedical data that maintain information about an individual’s health status and health care. Applications of deep learning to biomedical informatics research mainly focus on how to leverage EHR data for clinical decision support.

Medical images are one of the important resources stored in EHR. Following the traditional routine, researchers train deep learning models for feature representations and apply the pre-trained features to high-level tasks, such as classification, detection, and segmentation. Here are a few examples: for classification of tumor architecture, learning the features of histopathology tumor images with a deep neural net could improve the classification accuracy.

Use of convolutional neural networks (CNNs) even on non-medical image could improve identification of different types of pathologies in chest x-ray images; CNNs are also used to learn hierarchical representations of images for segmentation of tibial cartilage in low field knee MRI scans; a unified deep learning framework is developed for feature representation and automatic prostate MRI segmentation. Those studies show that researchers have attempted to apply deep learning models to clinical radiology research to assist physicians (Applications 1 and 2).

Application 1: detect disease, and describe their contexts from the patient chest x-rays

They present a deep learning model to efficiently detect a disease from an image and annotate its contexts.

They employ a publicly available radiology dataset of chest x-rays and their reports, and use its image annotations to mine disease names to train convolutional neural networks (CNNs).

Figure 2. An example of chest x-ray image

Comprehensive image understanding requires more than single object classification, to advances in deep convolutional neural networks (CNNs), effectively learning to recognize the images with a large pool of hierarchical representations. Most recent work also adapts recurrent neural networks (RNNs), using the rich deep CNN features to generate image captions. Learning from medical image text reports and generating annotations that describe diseases and their contexts have been very limited. Providing a description of a medical image's content similar to what a radiologist would describe could have a great impact. A person can better understand a disease in an image if it is presented with its context, e.g., where the disease is, how severe it is, and which organ is affected. Furthermore, a large collection of medical images can be automatically annotated with the disease context and the images can be retrieved based on their context, with queries such as "find me images with pulmonary disease in the upper right lobe.

A common challenge in medical image analysis is the data bias. In analogy to the previous works using ImageNet-trained CNN features for image encoding and RNNs to generate image captions, authors first train CNN models with one disease label per chest x-ray inferred from image annotations, such single disease labels do not fully account for the context of a disease. They employ the already trained RNNs to obtain the context of annotations, and recurrently use this to infer the image labels with contexts as attributes. Then they retrain CNNs with the obtained joint image/text contexts and generate annotations based on the new CNN features. With this recurrent cascade model, image/text contexts are taken into account for CNN training to ultimately generate better and more accurate image annotations.

The CNN-RNN based image caption generation approaches require a well-trained CNN to encode input images effectively. Unlike natural images that can simply be encoded by ImageNet trained CNNs, chest x-rays differ significantly from the ImageNet images. In order to train CNNs with chest x-ray images, they sample some frequent annotation patterns with less overlaps for each image, in order to assign image labels to each chest x-ray image and train with cross-entropy criteria.

They find 17 unique patterns of MeSH term combinations appearing in 30 or more cases, they use the aforementioned 17 unique disease annotation patterns to label the images and train CNNs. For this purpose, they adopt various regularization techniques to deal with the normal-vs.-diseased cases bias. For their default CNN model they chose the simple yet effective Network-In-Network (NIN) model because the model is small in size, fast to train, and achieves similar or better performance to the most commonly used AlexNet model . They then test whether their data can benefit from a more complex state-of-the-art CNN model, i.e. GoogLeNet from the 17 chosen disease annotation patterns, normal cases account for 71% of all images, well above the numbers of cases for the remaining 16 disease annotation patterns. They balance the number of samples for each case by augmenting the training images of the smaller cases where they randomly crop 224×224 size images from the original 256×256 size image.

it is difficult for a CNN to learn a good model to distinguish many diseased cases from normal cases which have many variations on their original samples, that normalizing via mini-batch statistics during training can serve as an effective regularization technique to improve the performance of a CNN model , Inspired by this and by the concept of Dropout, they regularize the normal-vs.-diseased cases bias via randomly dropping out an excessive proportion of normal cases compared to the frequent diseased pattern when sampling mini-batches. They then normalize according to the mini-batch statistics where each mini-batch consists of a balanced number of samples per disease case and a random selection of normal case samples.

They use recurrent neural networks (RNNs) to learn the annotation sequence given input image CNN embedding. Long Short-Term Memory: the LSTM has been successfully applied to speech recognition, sequence generation, machine translation, and, Gated Recurrent Unit: the GRU was successfully applied to machine translation.

In sampling they again initialize the RNN state vectors with the CNN image embedding (ht=1=CNN (I)). They then use the CNN prediction of the input image as the first word as the input to the RNN, to sample following sequences up to five words. As previously, images are normalized by the batch statistics before being fed to the CNN.

They evaluate the annotation generation on the BLEU score averaged over all of the images and their annotations in the training, validation, and test set. BLEU scores are a metric measuring a modified form of precision to compare n-gram words of generated and reference sentences. LSTM is easier to train, while GRU model yields better results with more carefully selected hyper-parameters 3. While they find it difficult to conclude which model is better, the GRU model seems to achieve higher scores on average.

Recurrent Cascade Model for Image Labeling with Joint Image/Text Context. CNN models are trained with disease labels only where the context of diseases are not considered, They use the already trained CNN and RNN to infer better image labels, integrating the contexts of the image annotations beyond just the name of the disease.

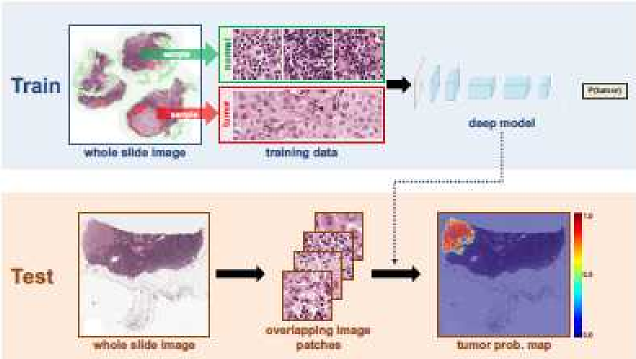
They present an effective framework to learn, detect disease, and describe their contexts from the patient chest x-rays and their accompanying radiology reports with Medical Subject Headings (MeSH) annotations. Furthermore, they introduce an approach to mine joint contexts from a collection of images and their accompanying text, by summarizing the CNN/RNN outputs and their states on each of the image/text instances. Higher performance on text generation is achieved on the test set if the joint image/text contexts are exploited to re-label the images and to train the proposed CNN/RNN framework subsequently.

Application 2: identifying metastatic breast cancer

The International Symposium on Biomedical Imaging (ISBI) held the Camelyon Grand Challenge 2016 (Camelyon16) to identify top performing computational image analysis systems for the task of automatically detecting metastatic breast cancer in digital whole slide images (WSIs) of sentinel lymph node biopsies.

The manual pathological review of sentinel lymph nodes is time-consuming and laborious, particularly in cases in which the lymph nodes are negative for cancer or contain only small foci of metastatic cancer. Many centers have implemented testing of sentinel lymph nodes with immunohistochemistry for pancytokeratins, which are proteins expressed on breast cancer cells and not normally present in lymph nodes, to improve the sensitivity of cancer metastasis detection. However, limitations of pancytokeratin immunohistochemistry testing of sentinel lymph nodes include: increased cost, increased time for slide preparation, and increased number of slides required for pathological review. Further, even with immunohistochemistry-stained slides, the identification of small cancer metastases can be tedious and inaccurate.

Figure 3. The framework of cancer metastases detection



They present a deep learning-based approach for the identification of cancer metastases from whole slide images of breast sentinel lymph nodes. Their approach uses millions of training patches to train a deep convolutional neural network to make patch-level predictions to discriminate tumor-patches from normal-patches. They then aggregate the patch-level predictions to create tumor probability heatmaps and perform post-processing over these heatmaps to make predictions for the slide-based classification task and the tumor-localization task. Combining the predictions of their deep learning system with a pathologist's interpretations produced a significant reduction in the pathologist's error rate.

They adopt a threshold based segmentation method to automatically detect the background region. In particular, they first transfer the original image from the RGB color space to the HSV color space, then the optimal threshold values in each channel are computed using the Otsu algorithm and the final mask images are generated by combining the masks from H and S channels. According to the detection results, the average percentage of background region per WSI is approximately 82%.

Their cancer metastasis detection framework consists of a patch-based classification stage and a heatmap-based post processing stage. During model training, the patch-based classification stage takes as input whole slide images and the ground truth image annotation, indicating the locations of regions of each WSI containing metastatic cancer. They train a supervised classification model to discriminate between these two classes of patches, and they embed all the prediction results into a heatmap image. In the heatmap-based post-processing stage, they use the tumor probability heatmap to compute the slide-based evaluation and lesion-based evaluation scores for each WSI.

During training, this stage uses as input 256 × 256 pixel patches from positive and negative regions of the WSIs and trains a classification model to discriminate between the positive and negative patches. They evaluated the performance of four well-known deep learning network architectures for this classification task: GoogLeNet, AlexNet, VGG16 and a face orientated deep network, the two deeper networks (GoogLeNet and VGG16) achieved the best patch-based classification performance. They adopt GoogLeNet as their deep network structure since it is generally faster and more stable than VGG16. The network structure of GoogLeNet consists of 27 layers in total and more than 6 million parameters. After generating tumor-probability heatmaps using GoogLeNet across the entire training dataset, they noted that a significant proportion of errors were due to false positive classification from histologic mimics of cancer. To improve model performance on these regions, they extract additional training examples from these difficult negative regions and retrain the model with a training set enriched for these hard negative patches.

There are two kinds evaluation in Camelyon16: Slide-based Evaluation and Lesion-based Evaluation. They won both of these two challenging tasks.

Slide-based Evaluation: the merits of the algorithms were assessed for discriminating between slides containing metastasis and normal slides. Receiver operating characteristic (ROC) analysis at the slide level was performed and the measure used for comparing the algorithms was area under the ROC curve (AUC). Lesion-based Evaluation: for this evaluation, free-response receiver operating characteristic (FROC) curve was used. The FROC curve is defined as the plot of sensitivity versus the average number of false-positives per image. The results of observations: first, the pathologist did not make any false positive predictions, second, when the average number of false positives is larger than 2, this indicates that there will be two false positive alert in each slide on average.

Application 3: predict healthcare-associated infections

Free-textual data are another important resource in EHR, including clinical notes, pathology notes, radiology reports, discharge summaries, etc. Those data contains massive valuable information provided by medical professionals. The problem is those data are not standardized, which makes it difficult for computers to process. Natural language processing (NLP) techniques are recently widely used in extracting information from clinical free-texts. HealthCare-associated infections(HAI) is defined as “an infection occurring in a patient in a hospital or other healthcare facility in whom the infection was not present or incubating at the time of admission.”

HAI detection is a main challenge in healthcare. Machine learning and natural language processing use electronic patient records to detect this. Electronic patient records usually contain abbreviation of concepts. For example Pat was op. two days ago, meaning the patient was operated two days ago. Deep learning architectures transform input data using several non-linear operations.

Artificial neural networks are a biologically inspired type of machine learning architecture. The basic building blocks of these networks are inspired by the biological neuron and are usually referred to as nodes. The nodes are connected to other nodes forming a network which can be trained by various methods to perform different tasks. Two different types of neural networks were used to train stacked sparse auto encoders and stacked restricted Boltzmann machines.

9. Stacked sparse auto encoders

A stacked sparse auto encoder is a neural network technology which consists of several sparse auto encoders. An auto encoder is a type of neural network that first encodes and then decodes input data. Auto encoders are trained to reproduce the data sent through them (Bengio, 2009). The cross entropy cost function was used for the auto encoders in this study. When the sparse auto encoders were stacked, each encoder was trained to reproduce the encoded data of the previous encoder.

This unsupervised training constituted the pretraining phase of the network. After the pretraining was completed only the first half of the auto encoders, responsible for the encoding, was used in the resulting network. A softmax classifier was appended to the last auto encoder and the network was trained in a supervised manner through back propagation.

10. Stacked restricted Boltzmann machines

A restricted Boltzmann machine (RBM) is a special case of the more general Boltzmann machine. An RBM consists of one layer of visible nodes and one layer of hidden nodes. Nodes in the hidden layer are only allowed to connect to nodes in the visible layer (Bengio, 2009). One notable difference between auto encoders and RBMs is that RBMs have binary activation of the hidden layer nodes during the training phase. In the implemented stacked RBM topology the RBMs were trained to reproduce data input on the visible layer through the means of persistent contrastive divergence. The stack of RBMs was created in a similar way to the previously described stack of auto encoders. Each RBM was trained on the hidden layer activations of the previous RBM. A softmax classifier was appended at the end of the network and was trained in a supervised manner on the hidden layer representations of the previous RBM. Unlike the stack of auto encoders, the whole network was not fine tuned through supervised training. Finally we can tell that The SRBM architecture was the most successful classifier in this study.

11. 3 ways data analytics can reduce total healthcare costs

Jennifer Goforth Gregory in www.ibmbigdatahub.com says: you want to provide the best care possible to each patient who walks through your hospital's doors, but you also need to keep cost of care low in order to keep those doors open and help patients for years to come. This can often be a tough balancing act for healthcare leaders. However, with the help of data analytics, your organization may be able to provide more personalized services to patients while reducing the total costs of care.

To date, these capabilities are largely underutilized: A KPMG survey of healthcare facilities found that only 10 percent of healthcare professionals use advanced data analytics tools with both analytic and predictive capabilities. By using analytics to reduce costs in ways that do not impact quality of care, you can use the savings to improve the service provided to each patient. Hospitals that do not use data analytics to reduce cost of care are effectively leaving money on the table.

The following is three ways that healthcare analysis and data analytics are currently helping medical providers find cost-saving opportunities.

12. Find and reduce inefficiencies

By using analytics, Beaufort Memorial Hospital in South Carolina determined that it could save an estimated $435,000 each year by discharging patients half a day early, according to U.S. News & World Report. The facility's IT department now analyzes 180 parameters and reports on key data points to plan prescriptions, follow-up visits, wheelchair transportation and room cleaning to reduce lag time between patient treatment steps. This has helped improve patient satisfaction and hospital costs.

Because Beaufort is typically at capacity, reducing the length of hospital stays by even just a few hours allows the hospital to better serve patients who are waiting for care. By setting small goals for the staff, such as assigning patients to beds in under 10 minutes, the hospital has seen an overall reduction in average stay length. Most importantly, staff support the process because the objectives are feasible.

13. Reduce emergency room visits

The Minnesota Department of Health recently found that 1.3 million unnecessary trips to hospitals and emergency rooms take place in the state each year, representing two out of every three ER patients and costing medical facilities $2 billion. Since this discovery, the state has started to use analytics to reduce preventable ER visits. For instance, the state pinpointed 50,000 residents that had at least four preventable ER trips in the past year due to chronic illnesses and worked with medical providers to ensure these individuals were receiving care in a primary care or community health setting. The state expects that these efforts to provide appropriate preventive care will reduce healthcare costs and overall hospital admissions since early treatment can decrease the severity of an illness.

14. Eliminate unnecessary testing

Healthcare Finance News reported that St. Louis Children's Hospital was able to reduce the number of $6,000 tests ordered for Dravet Syndrome, a rare form of epilepsy, by leveraging data analysis. These tests are often ordered because the standard of care for infants with seizures is to first check for chromosomal microarray, which is identified with the Dravet Syndrome test. However, Dr. Nephi Walton used data analytics to determine that 0 percent of Dravet Syndrome tests for this reason returned positive in the last five years. This finding caused the hospital to modify its standard of care and reduce the number of costly tests it was running.

It can be easy for hospitals to focus on sweeping changes for cost-saving measures, but often the easiest and most effective ways to save money are by reducing small resource waste issues that add up over time. These opportunities can often be hard to find on your own, but when you use data analytics, it becomes easier to identify ways to improve your processes and save significant money. The most important part is that not only are you improving your organization with these savings, you can also provide a higher level of care to the patients you treat every day.

15. Conclusion

Deep learning provides a solution to address the data analysis in massive volumes of input data. Deep learning approach does this operation automatically extracting complex data representations from very large collections of raw data that is generally unsupervised and un-categorized. Today big data and deep learning approach is used in many domain of healthcare, such as detecting disease x-rays image, identifying cancer, predicting disease with electronic health record (EHR), and cost reduction.

References

Alsheikh, M. A., Niyato, D., Lin, S., Tan, H. P. and Han, Z. (2016). Mobile big data analytics using deep learning and apache spark. *IEEE Network*, *30*(3), 22-29.

Chen, X. W. and Lin, X. (2014). Big data deep learning: challenges and perspectives. IEEE access, 2, 514-525.

Landset, S., Khoshgoftaar, T. M., Richter, A. N. and Hasanin, T. (2015). A survey of open source tools for machine learning with big data in the Hadoop ecosystem. *Journal of Big Data*, *2*(1), 24.

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R. and Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, *2*(1), 1.

Zhao, Y., MacKinnon, D. and Gallup, S. (2015). Big data and deep learning for understanding DoD data. *CrossTalk,*