Cloud-Based Patient Profile Analytics System for Monitoring Diabetes Mellitus

Shruthi M Kulkarni
M.tech Scholar
Department of Computer Science and Engineering
SIT, Tumkur - 572103, Karnataka, India

Dr.B Sathish Babu
Professor
Department of Computer Science and Engineering
SIT, Tumkur - 572103, Karnataka, India

Abstract—Healthcare systems are facing the challenge of information overload in caring for patients in a safe, affordable and high quality manner in a system with limited healthcare resources . To alleviate this problem, we develop a cloud-based patient profile analytics system for monitoring diabetes patient's data. Data analytics is nothing but drawing the conclusion about that information by examining the raw data. This healthcare system extracts the data from various sources and stores the information in a patient profile graph. The data consist of both structured and unstructured data. The patient profile graph provides a holistic view of the patient's information. The system can infer implicit information useful for clinical purposes and administration and also extracting the relevant information for performing predictive analytics. The predictive analysis can predict the risk of unplanned patient readmission and also the chances of getting the diabetes based on dietary plans and family history.

Keywords: Data Analytics, Healthcare Analytics, Cloud, Bigdata, Patient Profile, Predictive Analytics.

I. Introduction

Data analytics[1] is nothing but the communication and discovery of meaningful category of data. It draws the conclusion about the information by examining raw data. To quantify the performance, analytics relies on the simultaneous application of statistics, operational research and computer programming. The firms may commonly apply analytics to describe, predict and improve business performance. It focuses on inference, based on what is known by the researcher to derive conclusions. It describes everything from online analytical processing (OLAP) to CRM analytics. Modern data analytics is often supported by real-time data streams and are used by information dashboards. Analytics is technically the methodical use of data and related business introspection developed through applied analytical disciplines which drive the fact-based decision making for planning, learning, management and measurement. Analytics may be descriptive ,prescriptive or predictive. The healthcare payers and providers are under pressure to deliver better outcomes. The dynamics with respect to cost are changing, seeing the current environment, it is expected to become even more complex over the next several years. To enable the improved outcomes, the immense complexity confronting the healthcare industries requires smarter and improved decisions. Analytics provides the mechanism that can sort through a torrent of complexity of data and in turn help the healthcare organizations to deliver on these demands. Analytics is used to gain better insight which can help in demonstrating the value to achieve better outcomes, such as technologies and treatments . Information leading to insight can help informed and educated consumers become more accountable for their own health. This can improve efficiency and effectiveness. Analytics can assist discovery and exploration and it also aids in planning policy and programs by designing which in turn improves the operations and service deliveries, augment sustainability, palliate risk and provide a means for measuring and assessing fastidious organizational data.

Healthcare organizations[2] are increasingly using analytics to apply, unlock and consume new introspection from information. To meet the business challenges, the operational and clinical improvements are derived by using new analytical methods.Using the basic spreadsheets, reporting tools and application reporting modules from the traditional baseline for transaction monitoring , data analytics in healthcare fields is moving towards the model which will eventually incorporate predictive analytics and this enables the organizations to "see the future" create more personalized healthcare, allow dynamic fraud detection and predict patient behavior.

Cloud computing[3] has made it possible to obtain mainstream solution for data processing, distribution and storage, but moving large amounts of data in and out of the cloud presented an insuperable challenge for organizations with terabytes or petabytes of digital content that is big data[4]. Clouds helps in moving huge bytes of data at the speed ordered by businesses, they achieve transfer speeds that are unsuitable for such volumes by using fraction of available bandwidth, introducing unacceptable delays in moving huge bytes of data into and out of or within the cloud. It also helps in retrieving the large amount of data by any other suitable healthcare related users.
The proposed healthcare system adopts a repetitive process where in the system keeps interacting with the healthcare professionals or specialists as part of a feedback to gather, deduce, establish and enhance the self-learning knowledge base. The system consists of two main components: PROFILING and ANALYTICS. The profiling component extracts the data from various sources and stores the information in a patient profile graph. The patient profile graph provides the holistic and unified view of the patient's clinical data which simplifies the routine tasks performed by healthcare professionals. The analytics component analyzes the patient profile graphs to assume implicit information and extract relevant features for prediction tasks. Analytics utilizes the doctor's input and tag a small number of patients with the most informative data and it integrates with the analytics algorithm to provide expert hypotheses. This paper presents our results on using the healthcare system to predict the probability of patients being readmitted to hospital within few days of discharge. The prediction task which can also predict the risk of developing the diabetes is based on the patient's demographics, hospital utilization, ancestral medical history and dietary plans of the patients.

II. ORGANIZATION OF PAPER

Paper is organized as follows: section 3 includes the related works, section 4 includes the details of the proposed system along with the architecture, section 5 deals with the implementation details and section 6 draws the conclusions.

III. RELATED WORK

Diabetes, in one of the related works was divided into two subgroups based on the specific mechanisms causing the disease: diabetes associated with other diseases or conditions and diabetes in which genetic susceptibility is clarified at the DNA level. Smoking habits were reported to the Swedish National Diabetes Register (NDR) to study associations between smoking, micro albuminuria, glycemic control and identify trends associated with it. Specially in young females with type 1 diabetes and in middle-aged type 1 and type 2 diabetes patients, studies concluded that smoking habits in patients with diabetes were widespread,. The study recommended that these groups be targeted for smoking ceasing campaigns. Smoking habits were also associated with both poor glycemic control and microalbuminuria, which were found to be independent of other study characteristics such as weight loss goals among participants enrolled in an adapted Diabetes Prevention Program (DPP) were also investigated. In a real-world translation of the Diabetes prevention program, lifestyle intervention participants who achieved their weight loss goal were more likely to have monitored their dietary intake and frequency and were also more likely to have increased their physical activity markedly. To increase levels of physical activity and maintain dietary self-monitoring,these findings emphasized the importance of supporting participants in lifestyle interventions . A study has also been conducted to carry out the prediction analysis for treating hypertension using regression-based data mining techniques.

IV. PROPOSED SCHEME

The architecture of the healthcare system is illustrated in figure 1. The system takes clinical data from healthcare organizations and a medical knowledge base as input and provides integrative healthcare analytics for our target users such as doctors and administrators to analyze and predict the patient profiles.The healthcare system takes clinical data drawn from the multiple sources of patient data which can be either structured data containing patient’s demographics, lab test results, medical history etc., or unstructured data sources storing free-text doctor’s notes. To understand the unstructured data[10] , there are several natural language processing engines ,such as MedLee[7] and cTakes[11] and several medical dictionaries, such as unified Medical Language Systems(UMLS)[5].The system utilizes a well-known medical knowledge base UMLs[5] to interpret unstructured doctor’s notes that is identifying medical concepts and relationships between concepts.The healthcare system targets two kinds of users in healthcare organizations:
(i)administrators who manage clinical data (ii) medical professionals who query the data for managing the clinical care of patients.

The system provides various analytic outcomes to the users, such as, the system provides a holistic view of patient through patient profile graph which contains the comprehensive information of each patient the graphs interacts with the users to answer some of the typical questions asked by doctors like: list the patients who have hospital acquired infections or list all the patients who are taking specific treatments.The system answers questions related to quality of care such as the total number of patients readmitted to the hospital within 30days.The system supports various predictive tasks , such as identifying patients at high risk of developing diabetes in the near futurebased on their diets, ancestral medical history etc.,.
We discuss the various components of the system:

PATIENT PROFILING:The Profiling component constructs a profile graph for each patient from the clinical patient data that provides a holistic view of medical concepts and their relationships[13]. The graph contains the structured and unstructured data and various relationships between the identified concepts such as treatments, diagnoses etc.,. This component utilizes Natural Language Processing(NLP) engines to extract named entities called mentions. It then contrivances thecollective inferenceto simultaneously map mentions to their semantically matched concepts in the knowledge base and it also discovers the additional relationships.The component asks doctors to verify or attest mention-concept mappings and the identified relationships between the concepts to improve the accuracy of this process. The outcomes of the profiling component are:
1)Building patient profile graphs 2) localizing and improving our medical knowledge base.

HEALTHCARE ANALYTICS: The analytics component provides the healthcare analytics capabilities after constructing patient profile graphs. To perform the analytic tasks of our users such as, predicting whether the diabetic condition of a patient will be well-controlled or whether patients will be re-admitted within 30 days or identifying the patients who are at high risk of developing diabetes in the near future based-on their dietary plans, ancestral medical history etc.,. The following steps are taken by the component[13]. It first identifies the concepts or relationships in the patient profile graphs that are important to the particular analytic task. This identification process can be achieved by either applying automated feature selection techniques or features selected based on the input from doctors. The next step of the ANALYTICS component applies various analytics algorithms to the features and training data identified earlier. The analytics algorithms considered includes various classification, clustering and prediction techniques.The analytics component consists of three major steps as mentioned in Figure 2.

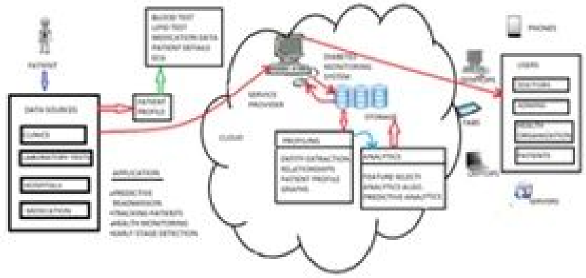

Figure 1: Patient Profile Analytics System Architecture

Feature selection: Substantially, all the features present in the patient profile graphs can be used as the features for the analytics tasks . Additionally, analytics can derive important and implicit features with expert input from the healthcare professionals or specialists. Analytics will verify the hypotheses and would revert back to the doctors to supportor reject their hypotheses with empirical evidence.

Training data labeling: In some of the prediction tasks, with well-defined class labels there is a lack of training

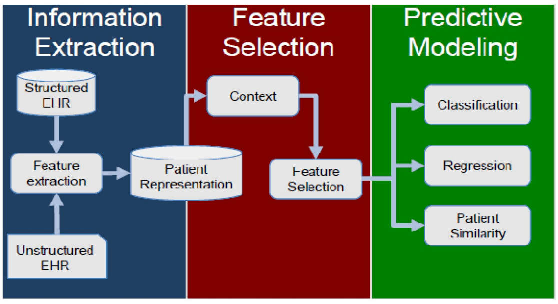

Figure 2: Analytics Steps

.

samples .In essence, what we need is a diverse set of labeled patients that somehow covers the whole data space as much as possible. For this purpose, analytics groups similar patient cases together and shows these groups to doctors. The purpose is to let the doctors freely select the groups or patients that they feel more comfortable to provide the labels. In addition, for each cluster, analytics presents only the features such that the patients in the cluster have similar values on these features. In this way, we avoid overwhelming the doctors with too much information on analytics might need to ask doctors to label patient profiles in several rounds. In particular, after obtaining the training data with annotated labels, analytics applies the machine learning algorithms and then continues to pick tuples that they have low confidence in predicting their class labels and ask the doctors for their input on the class labels of these tuples.

Analytics algorithms: Based on the derived features and training data, ANALYTICS exploits conventional analytics algorithms, such as classification, clustering and prediction to perform the various analytics tasks. In addition, the doctors might have some expert rules/heuristics for the analytics tasks.

PREDICTION: This section presents the healthcare system to predict the probability of the patients who have the high risk of developing diabetes-based on their diet habits and ancestral medical history and also predicting the probability of patients being readmitted to the hospital within few days of discharge[9]. The healthcare system uses the following features for the prediction tasks: patient's demographics such as age, gender and race, hospital utilization such as length of stay, previous hospitalizations and emergency visits, primary diagnosis and features derived from doctor's notes including laboratory results and past family medical history of their ancestors, as to whether they had diabetes or not and also based on their routine diet plans. The prediction of diabetes can be done on diet and family history, which is given in a detailed form below.

Diet: The major driving force behind intensifying diabetes epidemics worldwide is the excessive intake of calories, but diet quality also has its independent effects on diabetes. In the Nurse's Health Study (NHS)[12], we found that the important role in the development of diabetes quality is of fats and carbohydrates, independent of Body Mass Index and other risk factors. In particular, the trans fat and the higher dietary glycemic load (GL) are associated with increased diabetes risk, whereas there is a decreased risk in greater consumption of cereal fiber and polyunsaturated fat . In a meta-analysis, we found that a 2 serving per day increment in whole-grain intake was associated with a 21% lower risk of diabetes . Evidence also indicates that higher consumption of sugar-sweetened beverages would increase the risk of type 2 diabetes even after taking into account the effects of body weight. The analysis have found that the individuals who have the highest quantile of Sugar-sweetened beverages intake, most often 1–2 servings per day had a 26% greater risk of developing the disease.In addition to weight gain, several other mechanisms such as increased insulin demand, chronic inflammation and dyslipidemia may explain the adverse effects on cardiometabolic risk. Large quantities of rapidly absorbable carbohydrates such as sucrose result in a high dietary GL that leads to quick increases in blood glucose and insulin levels. A high GL diet, which increases insulin demand and may lead to pancreatic β-cell exhaustion in the long run which has been implicated in increased risk of diabetes . Fructose from high fructose corn syrup or any sugar may also play a role in this. It is preferentially metabolized to lipid in the liver, leading to insulin resistance. A study that compared the effects of consuming 25% of energy from fructose- or glucose-sweetened beverages showed similar kind of weight gain, but only the fructose group had a significant increase in visceral adiposity .Both animal and vegetable ghee , which are used for cooking in India and other countries, have an extremely high trans fatty acid content. A type of vegetable ghee called Dalda, a major source of edible oil in India, has 50% of trans fat level. Trans fat intake plays a role in the development of insulin resistance and chronic inflammation and is also associated with adverse cardio metabolic risk profiles and increased risk of heart disease.

Globalization and economic development have spurred nutrition transitions in many developing nations. This nutritional shift typically involves increased consumption of animal fat and energy-dense foods, decreased fiber and more frequent intake of fast foods have resulted in affecting the daily dietary plans of an individual which in turn is increasing the risk of developing the diabetes-based on the diets.

Family History: Family history is often considered as a risk factor for diabetes without even paying attention to the variable magnitude of risk-based on different familial risk profiles. The people with high and moderate familial risk of diabetes were more likely to report a diagnosis of diabetes than with the average risk persons in their families. Among risk factors examined, advanced age, non-white race, low education and income and self-reported obesity or weight gain were statistically significantly related to the diabetes risk and were evaluated to be the potential modifiers for the association between familial risk and diabetes. Thus the predictive analytics was used to predict whether diabetes was detected even based on the family history.

V. IMPLEMENTATION DETAILS

The supporting platforms are employed to support the afore mentioned Profiling and Analytics components. Firstly, the clinical data in the healthcare domain keeps growing dramatically.For instance, patients who are in intensive care unit are constantly beingmonitored, which would easily result in huge number of records of the patients. To address this scalability issue, we utilize EPIC[8], a flexible parallel processing framework, to support:

_ distributed data storage that effectively partitions the clinicaldata and stores them in multiple nodes.

_ scalable NLP processing and data analyticsthat involve various computation models, such as MapReduce model for entity extraction, Pregel model for graphical inference[6], deep learning for analytics, etc.

The second platform is for the interaction with our domain experts, i.e., the doctors or healthcare professionals. The platform is used to publish questions regarding the patients to doctors and collect their expertise suggestions. For instance, the proposed system can utilize the platform to leverage doctors to verify or support mention-concept mappings and concept relationships. Other similar examples may include asking the doctors tolabel training data and identifying key features for specific analytics tasks.

CONCLUSION

This paper presents a healthcare analytics system which allows point of care analytics for healthcare specialists who need to ask questions about the patients. The system extracts data from each patient from various data sources and stores them as information in a patient profile graph . The patient profile graph provides a holistic view and comprehensive information of patient's healthcare profile, which system can infer implicit information useful for administrative and clinical purposes and extract relevant features for performing predictive analytics. The predictive analytics helps in identifying the patients who have the high risk of developing the diabetes-based on the dietary plans of the patients and their ancestral diabetes medical history. At the core, the system keeps interacting with the healthcare professionals or specialists as part of a feedback loop to gather, assume, determine and enhance the self-learning knowledge base of the professionals.

REFERENCES

[1] http://searchdatamanagement.techtarget.com/definition/data-analytics

[2] http://www.ibm.com/smarterplanet/global/files/se__sv_se_healthcare__the_value_of_analytics_in_healthcare.pdf

[3] http://medicounts.com/2013-05-16-11-02-23/healthcare-analytics/hospital-optimization

[4] http://cloud.asperasoft.com/big-data-cloud/Ibm big data for healthcare.http://www.ibm.com.

[5] Unified medical language system. http://www.nlm.nih.gov/research/umls/.

[6] N. Allaudeen, J. L. Schnipper, E. J. Orav,R. M. Wachter, and A. R. Vidyarthi. Inability of providers to predict unplanned readmissions. J Gen Intern Med, 26(7):771–776.

[7] C. Friedman, P. O. Alderson, J. H. Austin, J. J.Cimino, and S. B. Johnson. A general natural- language text processor for clinical radiology. JAMIA, 1(2):161–174, 1994.

[8] D. Jiang, G. Chen, B. C. Ooi, K.-L. Tan, and S. Wu. epic: an extensible and scalable system for processing big data. PVLDB, 7(7):541–552, 2014.

[9] B. C. Ooi, K.-L. Tan, Q. T. Tran, J. W. L. Yip, G.Chen, Z. J. Ling, T. Nguyen, A. K. H. Tung, and M. Zhang. Contextual crowd intelligence.SIGKDD Explorations, 2014.

[10] S. Perera, C. A. Henson, K. Thirunarayan, A. P. Sheth, and S. Nair. Semantics driven approach for knowledge acquisition from emrs. IEEE J. Biomedical and Health Informatics, 18(2):515–524, 2014.

[11] G. K. Savova, J. J. Masanz, P. V. Ogren, J. Zheng, S. Sohn, K. K. Schuler, and C. G. Chute. Mayo clinical text analysis and knowledge extraction system (ctakes): architecture,component evaluation and applications. JAMIA,17(5):507–513, 2010. 1771

[12] http://care.diabetesjournals.org/

[13] Zheng Jye Ling, Quoc Trung Tran, Ju Fan, Gerald C.H.Koh, Thi Nguyen, Chuen Seng Tan, James W. L. Yip, Meihui Zhang.GEMINI: An Integrative Healthcare Analytics System.