AN ARTIFICIALLY INTELLIGENT SYSTEM TO PROVIDE HEALTHCARE SERVICES TO SOCIETY AND MEDICAL ASSISTANCE TO DOCTORS

Myrzakanurov Arman Sultanbekuly
Master student, International Information Technology University
Kazakhstan, Almaty

Duisebekova Kulanda Seitbekovna
Candidate of physical and mathematical science, Associate Professor, PhD
International Information Technology University
Kazakhstan, Almaty

Abstract. Since the last decade, number of applications of Artificial Intelligence in daily livelihood has drastically increased. This is primarily because of the inclusion of high-tech gadgets in our day-to-day lives. These gadgets provide high computational capabilities and geographical reach. These two features could be exerted to provide medical services to the society. This paper is based on a project which emphasizes on creating a software infrastructure which would provide healthcare services like diagnosis of diseases, advising medical tests to patients, providing medical prescription to patients by making use of personalized medicine problem solving algorithms etc., and providing medical assistance to doctors. This project, Medic, makes use of natural language processing, fuzzy logic, deep learning and a constantly evolving knowledge base to correctly diagnose diseases. It also provides various services to doctors which would help them while making decisions regarding any patient's medical treatment.

Keywords: Artificial Intelligence, Natural language processing, Image recognition, Convolution neural networks, Deep learning, Artificial neural networks.

I. INTRODUCTION

Having a healthy body is the biggest blessing we can possess. It is because of our good health, that we can overcome all the negativities around us and help the society in various possible ways. However, maintaining a healthy body is not a simple task since it requires large number of precautions to be taken. In addition to this, because of the fast Burden of Disease Study, it was found that more than 95% of the world's population is ill [4]. It clearly indicates the negligence shown by our society towards healthcare.

We can however take advantage of healthcare services offered by doctors and hospitals to retain our mental and physical health from illness. But, according to a research study published by One Medical Group that was gathered by Kelton Research in 2012, it was seen that majority of the respondents avoided doctors or hospitals despite being sick just because they believed it was a time consuming process. In the same research study it was seen that 45% of the respondents felt that visiting a doctor or hospital was a costly affair [4]. This project focuses on providing healthcare services to the people at their fingertips and that too at a low cost. Therefore Medic will not only save people's time, but also money. Hence, eventually more and more people will be attracted towards making use of this service and we will have a healthier society as a result of it.

The motto of Medic was to create cross-platform (major platforms such as smart-phones, tablets and personal computers) compatible software which would offer healthcare services to public. By making use of Medic, users can enter their health related problems along with some basic information and then the software will diagnose the disease and will provide a prescription for the same. In some of the cases Medic may ask its users to visit a hospital or a medical laboratory, if it thinks that certain medical tests are required to be taken by the users for proper diagnosis. If the software identifies any highly severe disease, then it will ask its users to visit a doctor for confirmation of that disease and proper physical treatment. Medic will provide basic information, medical history and previous prescriptions of the user to the doctor once the meeting is set up by it. This information would assist the doctor to cure the patient much faster than the normal approach. Medic will also offer patient records from all around the world to doctors, which would help them while making decisions.

II. RELATED WORK

In past, several efforts have already been taken in order to implement the knowledge of artificial intelligence in the field of medicine and healthcare (projects synonymous with [6] & [7]). There are computational devices and softwares which make use of AI algorithms to solve or assist in solving health problems. Out of these projects, some work at very detailed level by accepting features highly specific about certain disease or health problem. Gavin Robertson et al. [8] have written an article which describes a project falling into this category. They have accepted various features required for predicting blood glucose level. This project and other such similar projects are generally domain specific i.e., they can only be used to carry out tasks related to a limited number of health problems, but their results are found to be highly accurate. There are some other projects as well which accept primary symptoms from their users and use them to solve personalized medicine problem. Jamilu et al. [1] have described a project falling into this category in their research article. Such projects are much broader in scope. But lately, many researchers have doubted the accuracy of projects falling into this category, since some diseases need various parameters which can only be obtained from medical tests to be exactly diagnosed. Medic however, takes the positives from both the types of projects. It can diagnose nearly 1000 different diseases by making use of its default da-

tabase. This count increases when Medic comes across unknown scenarios; when it makes use of its self-learning capabilities and a web-crawler module to diagnose the unknown disease. This project can also get at detailed level by asking its users to undergo certain medical tests to obtain the data required to diagnose a disease which cannot be accurately diagnosed just by taking the primary symptoms into consideration.

III. ARCHITECHTURE

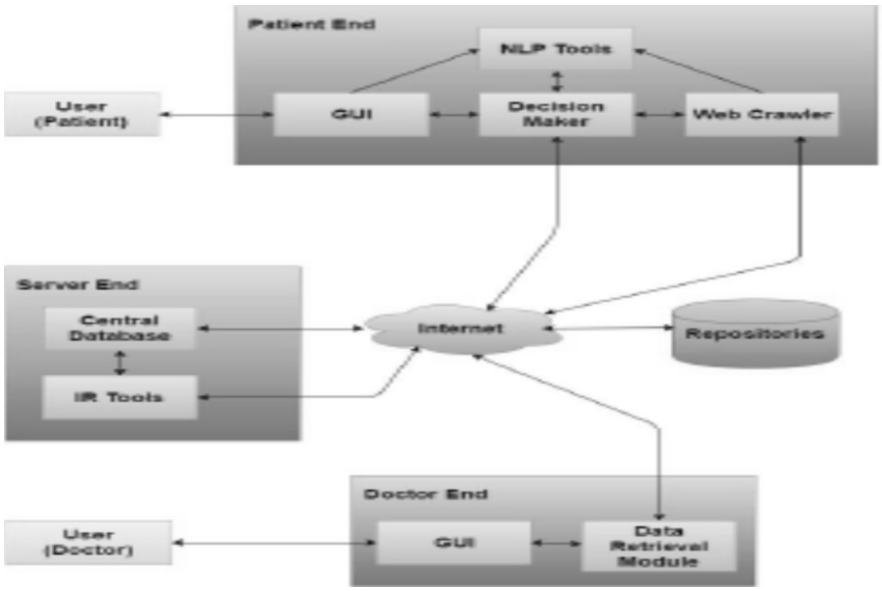

Fig. 1. Architecture diagram of Medic

Fig.1 shows the diagrammatic representation of the architecture of Medic. The project is architected such that it can implement self-learning procedures easily and that too without affecting any other module. The architecture can be divided into 3 parts based on the working of this project and its interaction with the users. These 3 parts are sequentially discussed in the subsequent sections.

A. Patient End

Patients provide their health concerns and get health assistance in return at this end. It has 5 modules.

1) Graphical User Interface (GUI): It is an interface which interacts with users to obtain their basic information and health related problems. Final results are shown over this module. It provides basic inputs to NLP module and decision maker module.

2) Natural language processing tools: These tools are required for transforming GUI or web crawler provided terms into the system known terms. It implements a sequence of procedures for carrying out this transformation. These procedures are, part-of-speech tagging, finding out unnecessary terms, neglecting unnecessary terms, building sets of terms by finding out synonyms of the terms and matching these sets with system known terms. Once this transformation is complete, the system can make sense out of the data provided by the user or the web crawler. NLP module of this project makes use Natural Language Toolkit [5] available in python to carry out intended tasks.

3) Web crawler: This module plays a vital role in providing self-learning abilities to this project. Web crawler is invoked only if the system comes across any unknown set of health problems. It is a simple python script which carries out search operations over a number of well-known healthcare web repositories. The results of these operations are generally data tuples, webpages and images related to user entered health problems. The data tuples are directly stored in the central database. The images are fed to the decision maker module along with some data for carrying out further operations. The webpages are fed to the NLP module to get sense out of the raw data. The decision maker module is provided with the results from the NLP module for updating the central database.

4) Decision maker module: This module can be imagined to be the brain of the patient end. It makes use of Fuzzy logic to make decisions. Once it receives inputs, first it decides whether to invoke the web crawler or not. This decision is made by fuzzy matching of the input with available data sets in the database. If the result is below a preset threshold, then web crawler is assigned a task to fetch information regarding the current input over the internet. Otherwise, the results are obtained from the database & existing convolution neural networks from IR module, aggregated and presented to the user.

B. Server End

The main objective behind having ‘Server End’ is centralization of data. Since all the high volume data and convolution neural networks are stored at this end, other two ends of Medic have very little memory requirement. This end consists of two modules.

1) Central Database: This database contains records of number of different diseases, their symptoms (stored as text fields as well as images) and drugs required to treat them. Medical history and previous prescriptions of every Medic user are also stored in the central database.

2) Image recognition tools: These tools are required for identifying the disease from visible symptoms provided by users. This module makes use of trained convolution neural networks for identifying the disease accurately. All the artificial neural networks of Medic are feedforward neural networks having multilayered architecture. GUI backend (at patient end) scales down the user entered image to some extent and then gives it to its decision maker module. The decision maker module makes a remote call to the IR module present at the server end.. Then it feeds the input image to the CNNs present in the IR module. It should be noted that all the convolution networks are stored on GPUs (Nvidia GTX-970) present at the server end. These CNNs then calculate results by simple matching. These results are provided to the decision maker module.

The IR module can be called by decision maker module for implementing the self learning procedure also. During this procedure, IR module is provided with an image or set of images along with the output they should yield. The IR module then first checks whether the input data is already present in the temporary memory(at server end). If a match is not found then it checks whether the temporary memory has enough space to accommodate input data. If enough space is available, then the data is stored in the temporary memory and the self-learning procedure halts. Otherwise, a new CNN (having architecture similar to the pre-existing CNN) is created in one of the GPUs. Then all the data stored in the temporary memory is extracted. This data is nothing but set of images and their corresponding outputs. Then this data is fed to the newly created CNN to carry out learning. This learning may take up to 2 to 4 hours.

C. Doctor End

This is the end at which doctors interact with Medic. Doctor end offers a GUI for presenting the requested data to the doctors. This end also has a data retrieval module, whose job is to provide user requested data, find out datasets stored in the database similar to the input dataset & publish the results.

VI. CONCLUSION

Vast developments in the field of technology can be implemented in medical science, in order to help the society. As far as Medic, this implementation heavily relies on artificial intelligence algorithms and techniques as discussed in this paper. Although, Medic is in its early phases of development; it can diagnose up to 1000 different diseases. This highlights the massive strengths possessed by AI techniques, points up extensive scope in the fields of AI to researchers and shows a glimpse of futuristic technology. Many forecasters think application of AI techniques could replace most of the jobs traditionally done by humans. However, from Medic we can proclaim that these technologies can assist us in our jobs and thereby make our daily lives easier and safer. Since Medic not only helps patients but also doctors, we can assert that Medic does not try to replace human doctors but rather tries to assist them. Finally, a successful implementation of AI techniques in the medical field would save many lives and help society being healthier.

References:

1. Jamilu Awwalu, Ali Garba Garba, Anahita Ghazvini, Rose Atuah.(December 2015). “Artificial intelligence in personalized medicine application of AI algorithms in solving personalized medicine problems”. International Journal of Computer Theory and Engineering, Vol. 7, No. 6, December 2015. [Online]. Available: http://www.ijcte.org/vol7/999-P030.pdf
2. Michael Nielsen. (2016, January). Neural networks and deep learning.[Online]. Available: http://www.neuralnetworksanddeeplearning.Com/Index.Html.
3. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey Hinton. "Imagenet Classification With Deep Convolutional Neural Networks". (2012).[Online].Available: http://www.cs.toronto.edu/~fritz/absps/imagenet.pdf
4. The 3 Most Common Reasons We Avoid Going to the Doctor.(2013, October 27). [Online]. Available: http://www.fool.com/investing/general/2013/10/27/the-3-most-common-reasons-we-avoid-going-to-the-do.aspx
5. Bird, Steven, Edward Loper and Ewan Klein (2009), Natural Language Processing with Python. O‟Reilly Media Inc
6. Anthony Farrugia, Dhiya Al-Jumeily, Mohammed Al-Jumaily, Abir Hussain, David Lamb. (December 2013). “Medical Diagnosis: are Artificial Intelligence systems able to diagnose the underlying causes of specific headaches?”. 2013 Sixth International Conference on Developments in eSystems Engineering.
7. Shih-Chung B. Lo, Shyh-Liang A. Lou, Jyh-Shyan Lin, Matthew T. Freedman, Minze V. Chien, Seong K. Mun. (December 1995). “Artificial Convolution Neural Network Techniques and Applications for Lung Nodule Detection”s. IEEE Transactions on Medical Imaging, Vol. 14, NO. 4, December 1995
8. Gavin Robertson, Eldon D. Lehmann, William Sandham, and David Hamilton, “Blood Glucose Prediction Using Artificial Neural Networks Trained with the AIDA Diabetes Simulator: A Proof-of-Concept Pilot Study,”Journal of Electrical and Computer Engineering, vol. 2011, Article ID 681786, 11 pages, 2011. doi:10.1155/2011/681786s

USE OF BIG-DATA IN HEALTHCARE AND LIFESCIENCE USING HADOOP TECHNOLOGIES

Arman Myrzakanurov Sultanbekuly
Master Student, International Information Technology University
Kazakhstan, Almaty

Abstract. Research and implementation of big data analytics is a booming area with huge latent to give valuable insights in the areas of healthcare and lifescience. While there are many challenges with implementation of big data analytics in its implementation. The major challenges of big data analytics are 3Vs (Volume, Velocity, and Variety), makes it difficult to implement it for clinical applications. However, the further research in the field of big data analytics has enabled its use in the field of healthcare and lifescience. The implementation of big data analytics helps in reducing the overall cost in the area of clinical research and helps in improving the accuracy of the predictions for diseases.

Keywords: Machine learning; Genomics; Lifescience; Public health care; Clinical trial; DNA.

I. INTRODUCTION

An era of open source datasets from reliable sources in healthcare is upcoming. There are already efforts made to digitize the health records by various medical institutes, and supported by various government organizations to make health data transparent. The data is curated and stored, which is further used by research institutes, pharmaceutical companies, and healthcare industries to provide useful insights.

The collected data is in form of big data as the volume is large and in addition to that the data has high complexity, variety and also velocity as it is collected from different sources continuously. This leads to the need for a technology which can address above stated issues and provide effective solution, Hadoop technologies including MapReduce, YARN comes in picture to achieve it. These technologies are in their early stages of their implementation, still it provides useful information to help the healthcare industry and reduces the cost.

For instance, researchers can perform big data analysis on the data to predict the treatments that are most effective for the given conditions, detecting diseases in early stage, reduce cardiac re-admittance rates, improvising clinical trials, effective cancer treatment and genomics and DNA fingerprinting.

Data scientists harness the power machine learning to detect autism and other diseases. Hospitals and patients can store their health record in the form of EHRs. It helps paper publishers to store medical data permanently.

The growth of big data analytics has led to opportunities for innovation, many companies are developing efficient solutions to achieve more accurate and substantial methods to reduce the soaring cost of healthcare.

The big data technologies are in early stage of its implementation and need further research and innovation to prove its reliability and effectiveness to stakeholders. This provides immense opportunities for researchers and industries to involve in early stage of its implementation. This paper will help to understand big data technologies such as Hadoop, MapReduce and YARN for implementing big data analysis software. This paper will also help in understanding possible solution in areas of medical treatment for healthcare and pharmaceutical industries.

II. BIG DATA AND HADOOP

A. Big-Data

Big data is a term for both structured and unstructured data. The data is of large volume with high complexity, which can not be processed by conventionally available applications (Excel, RDBMS, etc.).

The big data challenges include two things: 1) Data collection, storing data, and analysis of data. 2) Security and privacy of the information stored. The term big data extends its use to advanced methods of machine learning and deep learning to extract valuable information from data. With the use of advanced algorithms, accuracy in Big data analytics can be improved. This leads to confident decision making, and better decisions which result in the reduction of risks which in turn reduces costs and improved operational efficiency.

B. Hadoop 2.0

An open-source software framework provided by Apache, for storage and large-scale processing of datasets on clusters of commodity hardware. It is designed for scaling up from a single server to thousands of servers in a cluster with local computation and storage on every individual server. There are mainly three building blocks inside this run-time environment.

C. HDFS

HDFS stands for Hadoop Distributed File System, which comprises of two major components, namespaces and data nodes. The namespace service manages operations such as creation of files and directories over HDFS and modification of the same. The data node implements data node cluster management, block operations and replication of data stored over HDFS. HDFS Federation is the framework responsible for providing permanent, reliable and distributed storage. Intermediate data is not stored, rather the inputs and output data is stored. Other alternative storage solutions. For instance, Amazon uses the Simple Storage Service (S3). HDFS is an interesting technology in that it provides data distribution, replication, and automatic recovery in a user space file-system that is relatively easy to configure and, conceptually, easy to understand. However, its efficiency can be verified when MapReduce jobs are executed on data stored in HDFS.

III. BIG DATA SOLUTION TO HEALTH CARE AND LIFE SCIENCE

A. Early Disease Detection

Early disease detection is a way to detect early symptoms of a disease in an individual, data is collected to analyze the body profile including the blood test, ECG, pathological tests, X-Ray and scans. Big data ana-

lytics system can predict the disease and possible causes, in addition to that, it can predict the arrival of an outbreak when data is collected over a larger area, and it can go one step ahead and predict the disease and also the location of the carrier of a disease. This was done in the case of Ebola to neutralize its spread. Even in terms of personal care, body composition data can predict the side-effects of some medicines as well as the onset and spread of a disease. This is crucial information for doctors to prescribe proper medication to the patient.

B. Reduce Cardiac Re-Admittance Rates

Cardiac patients when admitted to the hospital, are well monitored and taken care by the hospital staff. After recovery, when they are back to home, there are chances that they may not maintain the proper self-monitoring instruction and follow a proper schedule of medications. This may lead to congestive heart failure which can cause a fluid retention. Researchers have found a small gain in weight due to phenomena of fluid retention.

C. Personalized Medicine and Treatment

Personalized treatment is a way to customize medical treatment based on creating a genetic body composition profile and lifestyle habits for an individual based on which a customized medical treatment plan is generated with the integration of continuous monitoring of effects of medication. Big Data analytics system can continuously collect data from various heterogeneous environment to monitor and computes patients' data and provide proper feedback. The medicine dose can be adapted based on the effect of medicines on an individual as predicted by the Big Data analytics system.

IV. CONCLUSION

Big data analytics is in early stage of its implementation in processing structured and unstructured data related to healthcare and lifescience. There are many startups and companies, which are using big data analytics to provide nnovative ways for data collection, analysis and predicting diseases.

References:

1. Jeffrey Dean and Sanjay Ghemawat,"MapReduce: Simplified Data Processing on Large Clusters", OSDI'04: Sixth Symposium on Operating System Design and Implementation, San Francisco, CA, December, (2004).
2. Harshawardhan S. Bhosale, Prof. Devendra P. Gadekar,"A Review Paper on Big Data and Hadoop", International Journal of Scientific and Research Publications, Volume 4, Issue 10, October 2014. 1. ISSN 2250-3153,(2014).
3. jStart: "How Big Data Analytics Reduced Medicaid Re-admissions." A jStart Case Study.(2012), http://www-01.ibm.com/software/ebusiness/jstart/portfolio/uncMedicaidCaseStudy.pdf.
4. Ikanow:"Data Analytics for Healthcare Creating Understanding from Big Data.", http://info.ikanow.com/Portals/163225/docs/data-analytics-for-healthcare.pdf.
5. Transforming Healthcare through Big Data Strategies for leveraging big data in the healthcare industry.(2013),http://ihealthtran.com/wordpress/2013/03/iht\%C2\%B2-releases-big-data-research-report-download-today/.