AI System for Prediction and Recommendation of Diabetes

Problem Statement

Evidence-based medicine is a powerful tool to help minimize treatment variation and unexpected costs. Large amount of healthcare data such as physician notes, medical history, medical prescription, lab and scan reports generated is useless until there is a proper method to process this data interactively in real - time. In this world filled with the latest technology, healthcare professionals feel more comfortable to utilize the social network to treat their patients effectively. To achieve this, an effective framework is needed which is capable of handling large amount of structured, unstructured and live streaming data about the patients from their social network activities. Healthcare Recommendation System (HRS)using machine learning can be developed to predict about the health condition by analyzing patient's life style, physical health factors, mental health factors and their social network activities.Forexample, on training the model with the age of women and diabetes condition helps to predict the chances of getting diabetes for new women patients without detailed diagnosis.

Background

In today's digital world people are prone to many health issues due to the sedentary life-style.The cost of medical treatments also keeps on increasing. An effective health care system is the one providing better personalized treatmentswith minimized cost. Medical expert systems are a branch of artificial intelligence that applies reasoning methods and domain specific knowledge to suggest recommendations like human experts. To enable reliable and fast decision-making process, medical expert knowledge needs to be stored as a knowledge based system(KBS). KBS alone is not sufficient to suggest reliable recommendations due to the limitations in updating expert rules based on the population studies and limited personalization.

Data driven approacheslike data mining and machine learning can be applied to extract insights from the heterogeneous data of the patients. It provides individual recommendations based on the past learning experience and the patterns extracted from clinical data. Combination of information retrieval and machine learning can be used for medical database classification. Various sources of knowledge for HRS are shown in figure 1.

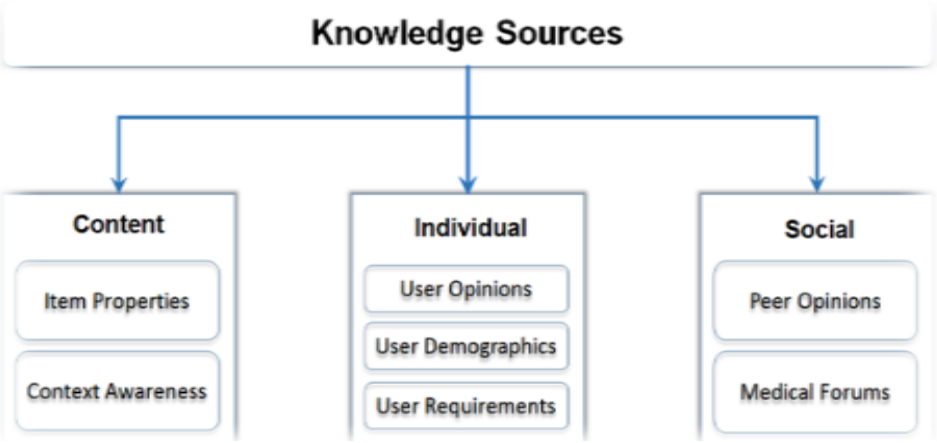

Figure 1: Knowledge sources for HRS

Methodology

Healthcare Recommender System (HRS) is designed for prediction of diseases and recommendation of treatment. It depends on a set of patient's case history, expert rules and social media data to train and build a model that can predict and recommend disease risk, diagnosis and alternative medicines.Predictions and recommendations need tobe approved by doctors. HRS system requires input information to generate predictions and recommendations.In this work diabetes data will be used as case study.

Step 1: Data collection and dataset preparation

Medical case history of diabetes patients will be stored which will contain information like blood sugar, bloodpressure, weight etc. Diagnosis data comprises of physician notes, lab results, and medications.The data records may have many attributes, values and doctor's diagnosis for each case. Diagnosis scale ranges from 1 to 5 based on the severity of the disease, 5-represents critical condition, 4-represents severe requires immediate treatment, and 3-represents moderate requires further investigation, 2-represents normal, 1-represents within control.Along with this, demographic data of active patient likename, age, location, education level, wearable device, lifestyle, food habits and type of connectivity will also be collected. A sample set of attributes of a patient are shown in figure 2.

Step 2: Developing a recommender system based on predictions using AI

Prediction is expressed as a numerical value that represents the disease risk diagnosis for future cases based on active patients. The rules for detecting various ranges for diabetes based on Fasting Plasma Glucose test (FPG), Casual Glucose Tolerance Test(CGTT)and Glycated Haemoglobintests (HBA1C)aregiven below:

Rule 1: No DiabetesRange

If FPG has a level between 70 and 100 mg/dL (3.9 and 5.6 mmol/L), then it indicates-no diabetes range.
If the blood glucose level below 125 mg/dL in CGTT, then it indicates- no diabetes range.
If HBA1Cvalue is below 97 mg/dL, then it indicates- no diabetes range.

Rule 2: Pre-diabetes Range

If FPG ranges from 100 mg/dl to 125 mg/dl and CGTT ranges from 140 mg/dl to 199 mg/d and HBA1C test values lie in range 97-154 mg/dL, it indicates pre-diabetes range.

Rule 3: Diabetes Range

If FPG is 126 mg/dl or more and CGTT is 200 mg/dl or more and HBA1C is greater than 180 mg/dL, it indicates diabetes.

The nest phase of recommendation which is expressed as the suggestion required by the users. For example, non-healthcare professional might be requiring alternative remedies for treating diabetes. The recommendation rules will be created by taking opinion from many doctors for different possible scenarios. Deep learning using CNN with auto-encoders or similar will be exploited for this task.

Step 3: Training and experimentation on datasets

Diabetes data set can be taken from KN specialtyclinic and downloaded from UCI repository.

Step 4: Deployment and analysis on real life scenario

The trained and tested recommender system will be developed in real-life scenario where historical medical records of diabetic patients will be collected from local hospitals.

Experimental Design

Dataset

Diabetes Data Set (https://archive.ics.uci.edu/ml/datasets/diabetes) will be used for experimentation.

Evaluation measures

Measures such as accuracy will be computed by comparing the prediction and recommendation from the doctors.

Software and Hardware Requirements

Deep learning libraries will be exploited for the development and experimentation of the project. Training will be conducted on NVIDIA GPUs for training the CNN mode

Attributes	Description	Values used
Age	Age of the user	Discrete Integer Values
Sex	Male or Female	Male or Female
BMI	Body Mass Index (Height to weight ratio)	Discrete Integer Values
Family History	Any family member of the subject is suffering/ was suffering from diabetes.	Yes or No
Smoking	Smoking habits of the user	Yes or No
Drinking	Drinking habits of the user	Yes or No
Lifestyle	Lifestyle of the user	Active, Moderate, Sedentary
Eating Habits	Food habits of the user	Healthy Foods, Junk foods
Frequent Urination	Urination habits of the user	Frequent or Normal
Increased Thirst	Urge to drink more than usual	Yes or No
Fatigue	Does the user feel fatigue often?	Yes or No
Blurred Vision	Do you have blurred vision?	Yes or No
Waist Size	Waist size of the user in inches	Discrete Integer Values
Gestational Diabetes	Do you have gestational diabetes?	Yes or No
Polycystic ovaries	Do you have polycystic ovaries?	Yes or No
Fasting Plasma Glucose	Values of Fasting Plasma Glucose	Discrete Integer Values
Casual Glucose Tolerance	Values of Random Glucose tolerance test	Discrete Integer Values

Figure 2: Sample attributes of a patient