

Application of Novel Deep Learning Algorithm in Disease Prediction from User's Health Related Information

Swati Y. Dugane¹ & Karuna G. Bagde²

¹Student, CSIT, HVPM's COET, Amravati

²Professor, CSIT, HVPM's COET, Amravati

^{1,2}SGB Amravati University, Maharashtra, INDIA

Abstract: Today, Human life is very busy. Sometimes we see that many people suffer from some disease but because of their time limit, they do not have time to meet doctors. Many people spent their time in searching internet. So if people know their symptoms that they suffer from. So they tell their symptoms to system. System then ask another question related to their disease, then user will answer it. After that system ask another question that is minimum 8 to 10 question about disease and finally display the disease name in which User suffer. So that before meet the doctors User have to know basic term about their disease and Get proper diagnosis. This is the aim of this paper. As well as after find their disease system also provide basic information about disease that is Cause, treatment, remedies as well as shows the photos of bacteria etc. System will also provide Government or private hospital located in user's area and also give information like hospital name, doctors name, timing, telephone number, in which days hospital is open or closed. This process is called deep learning. Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction

Key Words: Disease, Diagnosis, Remedies, Deep Learning

I) Introduction

With the easy availability of internet many people spend their time in searching basic and advanced health related information for their health related query. The information provided by search engine for eg. Google, is usually general, full of scientific terminologies and in scattered format. User may not be able to comprehend such haphazard information and usually leads to confusion or misinterpretation. Although a lot of work has already been done in this field to provide specific and targeted information to help the user, but every model has its own flaws in some or other way. To solve these problems, the proposed dissertation will develop a scheme based on user's health related data. It is nothing but deep

learning based on answers provided by user to the questions thrown by system for a primary health related symptom. The deep learning has two components. First is different medical signature from raw features, the raw features and signatures serve as input nodes in one layer and hidden nodes in the subsequent layer, respectively. The second learns the inter-relations between these two layers. In this we build sparsely connected deep learning architecture, which contain hidden layers. In this, user request will be compared with three different datasets. The datasets will be created using the information taken from authenticated medical literature (like well known textbooks, guidelines issued by authorized medical bodies and consultation with experts in the field). The system will save the medical information in the form of datasets and infer the possible diseases on the basis of symptoms and other health related data generated from health seeker.

II) Background

Luo et Al. [5] built a medical Web search engine called iMed, in 2008, which employed medical knowledge and an interactive questionnaire to help searchers form queries. S. Doan and H. Xu in , "Recognizing medication related entities in hospital discharge summaries using support vector machine," utilized SVM to recognize the medication related entities in hospital discharge summaries, and classified these atomic elements into pre-defined categories, such as treatments and conditions. which develop in 2010. In 2013, LiqiangNie and Tat-Seng presented a technique in "Beyond Text QA: Multimedia Answer Generation by Harvesting Web Information" [8].

III) Existing System

The D. A. Davis, N. V. Chawla, N. Blumm, N. Christakis, and A.-L. Barabasi presented the paper on "Predicting individual disease risk based on medical history," a novel system named CARE was designed in which combined collaborative filtering methods with clustering to predict each patient's greatest

disease risks based on their own medical history and those of similar patients. . Huang et al. proposed a re-ranking model for promoting diversity in medical search. Query-adaptive weighting methods that can dynamically aggregate and score various search results.

IV) Analysis and Discussion

1. The previous methodology used which unable to get exact answer because it contain did not contain complete information or data is not sufficient
2. Because of lots of datasets , performance level is low
3. Problem occurring in the association of datasets

V) Proposed Methodology

We are going to implement a scheme to finding the possible diseases according to questions of users. The medical data will store in the form of dataset. Every time the system responded to the user query according to the raw dataset. When user fires any query then the system accept that request and compare with the collected dataset. The dataset are nothing but the raw data. The main aim of this paper is to build a disease inference scheme that is able to automatically infer the possible diseases of the given questions in community based health services. The dataset forms base on authenticated medical book which form according to different categories such as diseases name, symptoms, precautions, descriptions etc. So to provide such facility of giving accurate to the users we will use the algorithm which is named as “Novel Deep Learning” algorithm. This algorithm is same as deep learning algorithm, but after adding some new features into it named as novel deep learning. We proposed and develop the scheme who studies the user information and health related data. In our application, the user request is compared with the different dataset. The main Objectives of this are:

A) To provide more accurate information to the user for their query:

In proposed system, every user will get more accurate information for their query. When compared with existing system, the proposed system gets more perfect result for users query. This will help the user for their disease prediction and the user can move to further procedure immediately as they get accurate disease inference.

B) To infer the possible diseases, given the questions of health seekers:

Proposed system identifies discriminant feature for specific disease. This means that there is no confusion in disease prediction according to users query. So that user can get accurate answer for their query. Distinguished answers will obtain to the user for the query. This means that for disease prediction proper query must be fired by health seekers. This will help the system to give proper prediction for user queries.

C) The system can remove the obstacles such as the vocabulary gap, incomplete information, correlated medical concepts:

There are various obstacles in the existing system like Vocabulary gap, incomplete information, correlated medical concepts. These obstacles are overcome in our proposed system. This will help the user to find the spellings of the diseases that they don't know. When we talk about the incomplete information, it means that only limited information is there in the system. Sometimes, this irritates the user and indirectly will affect on the reviews of the system.

D) System can able to find out the proper solution for users query:

All these objectives are key points in this proposed system, to get better results to the user so that every user gets satisfied results according to the query that user fired. So our system can be able to find out proper solution on the users query.

This paper aims to build a disease inference scheme that is able to automatically infer the possible diseases of the given questions in community-based health services. We first analyze and categorize the information needs of health seekers. It is worth emphasizing that large-scale data often leads to explosion of feature space in the lights of n-gram representations [8], [9], especially for the community generated inconsistent data. To avoid this problem, we utilize the medical terminologies to represent our data. Our scheme builds a novel deep learning model, comprising two components, as demonstrated in Figure 3.

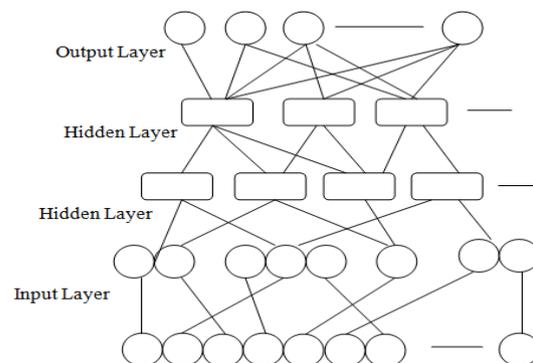


Figure1. Process of sparsely connected Deep learning

Extensive experiments on real-world dataset labelled by online doctors were conducted to validate our scheme. The main contributions of this work consist of 3 things:

- 1) This is the first work done on automatic disease inference in the community-based health services. Distinguished from the conventional sporadic efforts that generally focus on only a single or a few diseases based on the hospital generated records with structured fields, our scheme benefits from the volume of unstructured community generated data and it is capable of handling various kinds of diseases effectively.
- 2) The information needs of health seekers in the community-based health services get categorized and mines the signatures of their generated data.
- 3) It proposes a sparsely connected deep learning scheme to infer various kinds of diseases.

The Question and Answers (QA) session module is designed to perform the data preprocess. Tags are identified and categorized under the tag analysis. Features and signatures are identified under deep learning process.

1) Question and Answer Sessions:-

The Question and Answer (QA) data sets are collected from online health services. These sessions are done between user and system. The QA data values are transferred into the database. Questions, answers and tags are extracted from the datasets inside the database. The datasets are filled with category information.

2) Tag Analysis:-

Here tags are nothing but the query fired by the user to the system. The tags and associated disease information are identified in the tag analysis. Overlapped tag details are also updated with category information. Feature identification is performed in the tag analysis. Features and associated tag labels are updated into the database.

3) Deep Learning Process:-

Pseudo labeled data are analyzed in the learning process. Signatures get identified from the raw data under the learning process. Input layers and hidden layers are updated with features and signatures. The layers are used in the inference identification process. Raw data is nothing but data which we get by different sites like Medline Plus, WebMD, etc. In this way the complete process is get done by the system with the help of these three modules.

The Algorithm i.e. Novel deep learning algorithm is one deep learning algorithm. The difference between Shallow learning and deep learning is “shallow” machine learning, which is often based on user having some prior knowledge which specific features of the input may help in disambiguating the correct answer. The emphasis in shallow learning is not always on feature engineering and selection while in deep learning the emphasis is on defining the most useful computational graph topology and optimizing parameters correctly. So novel deep learning algorithm that we used here is:

Novel Deep Learning Algorithm

1. Start
2. Take query input from the user
3. Divide the data in different tokens.
 Input={i1,i2,i3.....in}
 Tokens={i1},{i2}.....{in}
 Dataset={d1.....dn}
4. Layer specification
 While(Layer)
 {
 Result R={in}U{d1.....dn}
 }
5. Formally, a one-hidden-layer is a function $f : R^D \rightarrow R^L$, where D is the size of input vector x and L is the size of the output vector $f(x)$, such that, in matrix notation:

$$f(x) = G(b^{(2)} + W^{(2)}(s(b^{(1)} + W^{(1)}x))),$$
6. with bias vectors $b^{(1)}, b^{(2)}$; weight matrices $W^{(1)}, W^{(2)}$ and activation functions G and s .

$$W^{(1)} \in R^{D \times D_h}$$
 is the weight matrix connecting the input vector to the hidden layer
7. Finally Different layer processing the result R given to user.

VI)System Requirements:-

Hardware Requirements :-

Processor: - Dual Core or more.

RAM:- 1GB

Hard Disk: - 40 GB

Software Requirements :-

Operating System : - windows 7, windows 8, windows 10,

Web Browser :- IE6 or upwards, Google Chrome, Mozilla Firefox

Web Server :- IIS 7.0

Development IDE Tools :- Microsoft Visual Studio

Database / Back End: - Microsoft SQL Server

biomedical information retrieval,” in Proceedings of the International ACM SIGIR Conference, 2009.

V. Conclusion

Numerous techniques are developed in the previous years like WebMD, MedlinePlus etc. The wide scope of this technique is presented in this paper. The accuracy level is achieved without huge time consuming unlike in previous works. The proposed system will help user to narrow down the list of differential diagnoses based on the symptoms, targeted medical history retrieved from the user and reports of some basic, noninvasive investigations. The ability of the system to arrive at or nearby a precise condition will certainly help the user to alleviate the curiosity as well as anxiety generated because of the symptoms. As the user will get basic first-hand knowledge of the medical condition, he/she will be able to make learned choices in dealing the condition during further workup in future. The proposed system will save the time as well as money, as it will be easy to approach the best medical facility in the locality as per user's affordability. As the system will also direct the user to consult the best suited healthcare facility for further course of action, it will add to the confidence, convenience and satisfaction of the user.

References

- [1] S. Doan and H. Xu, “Recognizing medication related entities in hospital discharge summaries using support vector machine,” in Proceedings of the International Conference on Computational Linguistics, 2010.
- [2] B. Koopman, P. Bruza, L. Sitbon, and M. Lawley, “Evaluating medical information retrieval,” in Proceedings of the International ACM SIGIR Conference, 2011
- [3] Liqiang Nie and Tat-Seng Chua “Beyond Text QA: Multimedia Answer Generation by Harvesting Web Information” IEEE Transactions on Multimedia, 2013.
- [4].D. A. Davis, N. V. Chawla, N. Blumm, N. Christakis, and A.-L. Barabasi, “Predicting individual disease risk based on medical history,” in The International Conference on Information and Knowledge Management, 2008.
- [5].X. Huang and Q. Hu, “A bayesian learning approach to promoting diversity in ranking for