MINING HEALTH RELATED DATA BY IMPROVISED MACHINE INTENSE LEARNING TECHNIQUE

J.Indhumathi and J.Sudha

Department of Computer Science and Engineering,

A.V.C College of Engineering.

ABSTRACT

Deep learning is a branch of machine learning based on a set of algorithms that attempt to model high-level abstractions in data by using model architectures, with complex structures. The main objective of this project is to do disease inference based on the symptoms that the health seekers give. It involves collection of medical related datasets and then doing signature modelling. Signature modelling here refers to inter dependent medical attributes. Through feature extraction, the given input is transformed to reduced set of features. A novel deep learning scheme to infer the possible diseases given the questions of health seekers. The hidden nodes serve as raw features for the more abstract signature mining. Various combinations of user given query are composed by including some set of conjunctions. Then that output is again mined as same symptom can occur for different disease. A machine learning approach to infer the possible diseases based on the symptoms the health seekers gave and on the trained data sets. The input will be the raw data, the output will be predicted disease and the intermediate layers will be hidden from the data.

INTRODUCTION

Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviours, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. The most commonly used techniques in data mining are Artificial Neural Networks, Decision trees, Genetic Algorithms, Nearest Neighbour method, Rule Induction.

Deep learning approach has been used for getting accurate disease prediction. Deep learning is a class of machine learning algorithms that use a cascade of many layers of nonlinear processing units for feature extraction and transformation. Each successive layer uses the output from the previous layer as input. The algorithms may be supervised or unsupervised and applications include pattern analysis (unsupervised) and classification (supervised).are based on the (unsupervised) learning of multiple levels of features or representations of the data. Higher level features are derived from lower level features to form a hierarchical representation are part of the broader machine learning field of learning representations of data. Learn multiple levels of

representations that correspond to different levels of abstraction; the levels form a hierarchy of concepts.

The current prevailing online health resources can be roughly categorized into two categories. One is the reputable portals run by official sectors, renowned organizations, or other professional health providers. The project disseminating up-to-date health information by releasing the most accurate, well-structured, and formally presented health knowledge on various topics. The time could vary from hours to days. Second, doctors have to cope with an ever-expanding workload, which leads to decreased enthusiasm and efficiency. From the training data set collected, our proposed approach involves extracting and mining valid data the health seekers need and finally presenting it to the user as a disease inference report.

PROPOSED DESIGN

The Proposed work involves mining of medical related data using Intense Machine Learning. It involves intermediate hidden layer construction. Each hidden layer contains partial output and it is refined until exact inferences are captured. The system involves mining signatures. The signature in this context refers to medical attributes. For example if the user gives the symptom as blurry vision, it is not only the symptom of diabetes .May be the user have the defect by birth. It involves mining the dataset with user given query and produces some partial inferences. In this phase the basic symptoms for each disease is collected. In medical field, the same thing can be called with different terminologies. All those things will be grouped under single criteria. Word net database, which has the collection all words of English along with its synonyms and definition, is used for doing feature extraction. The Level 0 gives all the synonyms for each word given in the query. In the signature modelling phase level 0 output is produced which is given as the input for this phase. Each word is checked with possible disease symptoms and the unnecessary words are excluded. Various combinations of user given query are composed by including some set of conjunctions. Then checked with the collected symptoms. Then that output is again mined as same symptom can occur for different disease. The mining is done repeatedly until it matches with threshold value. The disease name which matches with most of the symptoms can be presented to the user as the disease inference report. The raw feature given as input and the output layer consists of possible disease inference. The Intermediate layers are the hidden ones which involves partial output. Here the hidden layers are formed repeatedly until it satisfies the threshold values.

DISEASE INFERENCE

As aforementioned, vocabulary gap, incomplete information, inter-dependent medical attributes and limited ground truth have greatly hindered the performance of classic shallow machine learning approaches. To tackle these problems, project propose a novel deep learning scheme to infer the possible diseases given the questions of health seekers. Compared to shallow learning, deep learning has several advantages. First, it is able to learn representative and scalable features from other disease types. Take the lung cancer inference learning as an example. When building its classifier, the training data can be liver cancer or other disease samples rather than strictly constrained to lung cancer. This addresses the limited ground truth and necessity of disease-aware feature extraction. Second, inherited from its deep architectures, it repeatedly learns the more abstract compact patterns layer by layer. This enables the system to mine the underlying connections among medical attributes. Third, deep learning can seamlessly integrate signatures as hidden nodes. As analysed

previously, signatures infer the incomplete information. Most importantly, with deep learning, each data instance will be ultimately represented by a mixture of very high-level abstract patterns, which are semantic descriptors and thus are more robust of data inconsistency caused by vocabulary gap.

MACHINE LEARNING ALGORITHM

Input query

Let ar be an array, ar[]=Split up of query at the position of spaces

Construction of unnecessary words array

Exclusion of unnecessary words from the

input WordNet:

Set property of wordnetdatabase path to system path

Generate instance of wordnet

database Let res be resultant array,

For each word w, in ar

{

Synsetss[]=Generate synset for

w For each synset s in ss,

{

Res=getWordDefinitions();

}

}

Return res;

Compute possible combinations of

word Let al be arraylist;

For each word w, in wordary,

{

Ar[]=Get synonym for w;

For (each item I,inar)

{

Al.add(W+" "+I);

Al.add(W+" of "+I);

Al.add(W+" on "+I);

Al.add(W+" and"+I);

Al.add(W+"in "+I);

}

}

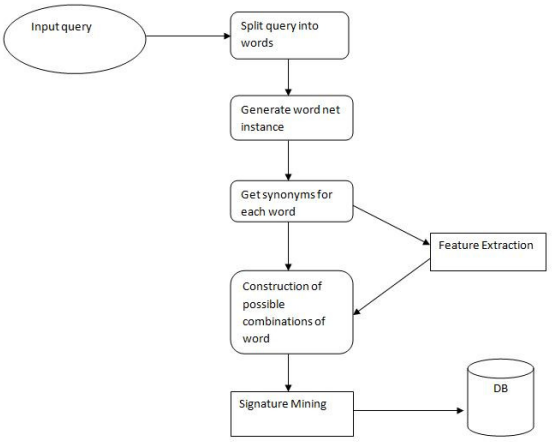
Iterate the list

Check each item in the list with the collected data

Get maximum matched disease name

Generate report

ARCHITECTURE DIAGRAM



CONCLUSION

Thus the proposed system involves an efficient intense machine learning approach for mining health related data. It involves considering discriminated features also for mining a particular disease. The hidden layers between the input and output layers are incrementally increased based on the accuracy

factor. The system can be further extended to compute adverbial and adjective form of the given query for further pruning of mining medical related data and getting accurate results.

FUTURE ENHANCEMENT

The future work involves mining of medical related data using Intense Machine Learning. It involves intermediate hidden layer construction. Each hidden layer contains partial output and it is refined until exact inferences are captured. The system involves mining signatures. The signature in this context refers to medical attributes. For example if the user gives the symptom as blurry vision, it is not only the symptom of diabetes .May be the user have the defect by birth. It involves mining the dataset with user given query and produces some partial inferences. In this phase the basic symptoms for each disease is collected. In medical field, the same thing can be called with different terminologies. All those things will be grouped under single criteria. Word net database, which has the collection all words of English along with its synonyms and definition, is used for doing feature extraction. The Level 0 gives all the synonyms for each word given in the query. In the signature modelling phase level 0 output is produced which is given as the input for this phase. Each word is checked with possible disease symptoms and the unnecessary words are excluded. Various combinations of user given query are composed by including some set of conjunctions. They are then checked with the collected symptoms. Then that output is again mined as same symptom can occur for different disease. The mining is done repeatedly until it matches with threshold value. The disease name which matches with most of the symptoms can be presented to the user as the disease inference report. The raw feature given as input and the output layer consists of possible disease inference. The Intermediate layers are the hidden ones which involves partial output.

Here the hidden layers are formed repeatedly until it satisfies the threshold values.

REFERENCES

[1] D. Zhang and W. S. Lee-Extracting key-substring-group featuresfor text classification in Proc. 12th ACM SIGKDD Conf. Knowl. Discovery Data Mining, 2006, pp. 474–483.

[2] G. Luo and C. Tang-On iterative intelligent medical search in Proc. 31st Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2008, pp. 3– 10.

[3] I. Batal, L. Sacchi, R. Bellazzi, and M. Hauskrecht-A temporal abstraction framework for classifying clinical temporal data in Proc. Amer. Med. Informat. Assoc., 2008, pp. 29–33. [4] J. Zhou, L. Yuan, J. Liu, and J. Ye-A multi-task learning formulation for predicting disease progression in Proc. 17th ACM SIGKDD Conf. Knowl. Discovery Data Mining, 2011, pp. 814– 822.

[5] J. Zhou, J. Liu, V. A. Narayan, and J. Ye-Modeling disease progression via fused sparse group lasso in Proc. ACM SIGKDD Conf. Knowl. Discovery Data Mining, 2012, pp. 1095–1103.

[6] L. Nie, M. Akbari, T. Li, and T.-S. Chua-A joint local-global approach for medical terminology assignment in Proc. Int. ACM SIGIR Workshop, 2014, pp. 24–27.

[7] L. Nie, Y.-L. Zhao, M. Akbari, J. Shen, and T.-S-Bridging the vocabulary gap between health seekers and healthcare knowledge Chua IEEE Trans. Knowl. Data Eng., vol. 27, no. 2, pp. 396–409, Jun. 2014.

[8] R. Fakoor, F. Ladhak, A. Nazi, and M. Huber-Using deep learning to enhance cancer diagnosis and classification- presented at the Int. Conf. Mach. Learn., Atlanta, GA, USA, 2013.

[9] T. C. Zhou, M. R. Lyu, and I. King-A classification-based approach to question routing in community question answering in Proc. 21st Int. World Wide Web Conf., 2012, pp. 783–790.

[10] X. Huang and Q. Hu- A Bayesian learning approach to promoting diversity in ranking for biomedical information retrieval in Proc. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2009, pp. 307–314.