2016

Towards developing an intelligent agent to assist in patient diagnosis using neural networks on unstructured patient clinical notes: Initial analysis and models

Ma. Regina Justina E. Estuar
Ateneo de Manila University

Christian E. Pulmano
Ateneo de Manila University

Recommended Citation

Christian E. Pulmano, Ma. Regina Justina E. Estuar, Towards Developing an Intelligent Agent to Assist in Patient Diagnosis Using Neural Networks on Unstructured Patient Clinical Notes: Initial Analysis and Models, Procedia Computer Science, Volume 100, 2016, Pages 263-270, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2016.09.153.

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 100 (2016) 263 – 270

Procedia Computer Science

Conference on ENTERprise Information Systems / International Conference on Project MANagement / Conference on Health and Social Care Information Systems and Technologies, CENTERIS / ProjMAN / HCist 2016, October 5-7, 2016

Towards developing an intelligent agent to assist in patient diagnosis using neural networks on unstructured patient clinical notes: Initial analysis and models

Christian E. Pulmano[a]*, Ma. Regina Justina E. Estuar[b]

[a]*Department of Information Systems and Computer Science, Ateneo de Manila University, 1108 Quezon City, Philippines*
[b]*Ateneo Java Wireless Competency Center, Ateneo de Manila University, 1108 Quezon City, Philippines*

Abstract

Technological advances in information-communication technologies in the health ecosystem have allowed for the recording and consumption of massive amounts of structured and unstructured health data. In developing countries, the use of Electronic Medical Records (EMR) is necessary to address the need for efficient delivery of services and informed decision-making, especially at the local level where health facilities and practitioners may be lacking. Text mining is a variation of data mining that tries to extract non-trivial information and knowledge from unstructured text. This study aims to determine the feasibility of integrating an intelligent agent within EMRs for automatic diagnosis prediction based on the unstructured clinical notes. A Multilayer Feed-Forward Neural Network with Back Propagation training was implemented for classification. The two neural network models predicted hypertension against similar diagnoses with 11.52% and 10.53% percent errors but predicted with 54.01% and 64.82% percent errors when used on a group of similar diagnoses. Further development is needed for prediction of diagnoses with common symptoms and related diagnoses. The results still prove, however, that unstructured data possesses value beneficial for clinical decision support. If further analyzed with structured data, a more accurate intelligent agent may be explored.

Peer-review under responsibility of the organizing committee of CENTERIS 2016

Keywords: Text mining; neural networks; intelligent agents

* Corresponding author. Tel.: +63998564097.
E-mail address: christianpulmano@gmail.com

Peer-review under responsibility of the organizing committee of CENTERIS 2016
doi:10.1016/j.procs.2016.09.153

1. Introduction

The implementation of various health information systems allows health stakeholders to easily capture patient health information. Healthcare leaders even find themselves overwhelmed with data, but still unable to identify those that will be helpful for making the right decisions[1]. The challenge at hand is being able to extract information from data and use it to a more informed decision-making and planning. Current technologies in the health ecosystem have already allowed several patient records to be combined and analyze[2] so that valuable and reliable information may be discovered. With the use of mathematical and algorithmic-based processing of data resources, data mining and analytics methods will help uncover information hidden within health data and develop descriptive, predictive and prescriptive models for deriving insights from data[3].

Data mining can be defined as the "process of finding previously unknown patterns and trends and using it to build predictive models[4]". Various techniques, such as clustering, association, and classification, are already available for data mining applications in healthcare. These techniques help transform the overwhelming amount of data into useful and actionable information[1]. Data mining is becoming increasingly essential in an evidence-based decision making process[4]. Unfortunately, data mining methods require a highly structured format of data[5]. Text mining, on the other hand, is a variation of data mining that tries to extract interesting and non-trivial information and knowledge from unstructured text. It is also sometimes referred to as Intelligent Text Analysis, Text Data Mining or Knowledge-Discovery in Text (KDT). While there are more benefits in recording structured data, there is still a potential value in also analyzing unstructured inputs in patient clinical encounters[6].

This study aims to determine the feasibility of integrating an intelligent agent within Electronic Medical Records (EMR) that would allow for the prediction of patient diagnosis based on clinical text data. Predicting diagnosis based on symptoms is an important problem, especially in the context of developing regions, where there are still a lot of shortcomings in the health information systems. The study focuses on the possibility of developing an artificially intelligent agent that implements continuous and automatic learning, and pattern detection from previously stored data. A main application of the intelligent agent is in the consultation workflow within EMRs wherein the intelligent agent will suggest possible diagnoses after the medical doctor inputs the signs and symptoms of the patient. Another important application is its ability to aid in the use of the International Classification of Diseases-10 (ICD-10) for standard diagnosis tagging.

This paper is organized as follows: in section 2 we present a review of previously conducted researches, and literature related to our study; in section 3 we identify the materials and methods used in the study; in section 4 we present the results and analysis of the initial models; and in last sections we state the conclusion and enumerate recommendations for further studies.

2. Related literature

Health information systems capture many types of health data, thus, a large amount of health data is collected and stored. This repository can offer great potentials in knowledge-seeking in the health and medical ecosystem. Data and text mining techniques can help draw out important insights and findings that could aid health stakeholders in decision-making processes.

2.1. Electronic medical records in the Philippines

Electronic Medical Records (EMR) are characterized as computerized legal medical records used within a facility[2]. EMRs can replace paper-based medical recording, provide a single point of access for data, automate report generation and track information trails, thus, improving the clinical healthcare delivery[7]. SHINE OS+ is an example of an EMR that is being used in selected sites in the Philippines. SHINE (Secured Health Information and Network Exchange) was launched in 2011 to address the need of the country for an integrated health management system[8]. The four main features of SHINE include four R`s: record, report, remind and refer. Among various EMRs being used in the Philippines are CHITS, WAH, SegRHIS, iClinicSys and for mobile-based platforms, eTABLET and eHATID. Though distinct, all EMRs are required to contain minimum standard datasets provided by the Department of Health (DOH).

2.2. Text mining and natural language processing in clinical data

Multiple studies have already utilized the analysis of unstructured text data present in electronic health records (EHR). One study used text analysis of clinical notes for automatic identification of heart failure (HF) diagnostic criteria. A natural language processing (NLP) engine was developed to identify Framingham HF diagnostic criteria from documented but un-synthesized health data. The NLP pipeline developed performs two levels of analysis, (1) annotates all affirmative and negative mentions of Framingham criteria using rule-based NLP system and, (2) label encounters whether the Framingham criterion is asserted, denied or unknown. It was concluded that the system was able to identify and label affirmations and denials of Framingham diagnostic criteria using primary care clinical notes[2].

2.3. Artificial neural networks in disease prediction

Artificial Neural Networks (ANNs) are "collections of neuron-like processing units with weighted connection between units[10]". It is a desirable tool for medical purposes because of its high parallelism, robustness, generalization and noise tolerance[11]. Several studies have already explored the use of ANNs in disease risk prediction. One study used fuzzy backpropagation neural networks for early diagnosis of Hypoxic Ischemic Encephalopathy in newborns[12]. Another study applied neural networks for diagnosis and survival prediction for colon cancer[11]. Both studies presented promising results in disease prediction which suggests further exploration in the computational potentials of ANNs in the health and medical field[11,12].

2.4. Agent-based systems for health decision support

In one study, an agent-based intelligent medical diagnosis system (AIMDS) was developed. It possesses the capacity to output appropriate medical prescriptions and food prohibitions given the physical signs and symptoms of the patient. It consists of three core modules: (1) sensing module, (2) reasoning module, and (3) medical knowledge module. The sensing module consists of sensors that capture signs and symptoms of the patient. The reasoning module was based on intuitionistic fuzzy set theory that finds the correlation measure of a set of signs and symptoms to related diagnoses. The knowledge module contains the field knowledge (i.e. symptom-disease match and disease-medicine match) required to support the reasoning module. Results of the developed agent-based systems demonstrated that it is feasible and practical to implement intelligent agents for healthcare decision-support. The agent was able to provide accurate diagnostic results which makes it a powerful tool for solving problems in medical care service[13].

The integration of agent-based systems for health decision support have already progressed in the field of electronic health. Such systems are able to create decisions given some circumstances based on already defined rules. Rules can be stored manually, via actual inputs and contributions from medical experts, and/or automatically, via extraction from data using machine-learning algorithms[14]. The most important factor in the success of intelligent agents is its ability to learn and update its rules based on new inputs. The implementation of health standards such as ICD codes and HL7 makes agent-based system even more effective[14,13].

3. Materials and methods

This section describes the materials and methods used in the study.

3.1. Source of data

Anonymized and deidentified data was extracted from the SHINE OS+ EMR upon obtaining approval from the data administrators. The final dataset used contained the following fields: complaint (TEXT), complaint history (TEXT), physical exam (TEXT) and diagnosis (TEXT). These fields are captured during the consultation and diagnosis process. Enumerated fields are all stored as text in the database via free-text input fields in SHINE OS+.

The SHINE OS+ database consisted of 81,516 individual patient records and 62,261 consultation data, gathered from the previous and current version of SHINE OS+ starting from January 2011 to January 2015. Data from SHINE OS+ was captured in different regions of the Philippines, both from private and public practitioners. Multiple

consultation records may belong to only one patient. The final dataset used for the study was composed of those with diagnosis enumerated in Table 1. For some parts of this study, `Pneumonia' and `Bronchitis' were grouped together as `Lower Respiratory Tract Infection' (LRTI).

Table 1. Diagnosis used and number of records.

Diagnosis	Records
Asthma	765
Bronchitis	2,953
Influenza	339
Pneumonia	2,037
Tuberculosis	1,715
Upper Respiratory Tract Infection (URTI)	6,775
Hypertension	1,773
Total	16,357

3.2. Process

Text analysis and classification were conducted using the following tasks: (1) Data Retrieval, (2) Pre-processing, (3) Processing and Analysis.

Data Retrieval. Anonymized and deidentified data were retrieved from the SHINE OS+ database and then exported to a local database. As previously mentioned, this study focused on text data in the consultation records. The attributes include (1) complaints, (2) complaint history, (3) physical exam and (4) diagnosis. All fields used for this study were stored as text fields. The diagnosis attribute was considered as the record label.

Pre-processing. Pre-processing includes vector creation of tokens from the complaint, complaint history and physical exam attributes. The text fields for each consultation record were tokenized and combined in one single vector. The Term Frequency-Inverse Document Frequency (TF-IDF) was used as measure for the attributes in the vector. Some tokens were removed based on the predefined list of stopwords for English. The tokens were further filtered by accepting only those with length greater or equal to three characters. This removed tokens such as “aa”, “ab” that were observed to be unnecessary for the study.

Processing and Analysis. Text mining techniques were applied to patient clinical text data (i.e. complaint, complaint history, physical exam) to conduct the analysis. A Python neural network classifier was designed using functionalities from the PyBrain package[15]. Processing and analysis accepts the filtered vector of tokens as input data for the classification model. The classification process uses a Multilayer Feed-Forward Neural Network with Back Propagation training.

Classification. In this study, the aim is to be able to identify if text data found in the patient clinical encounter notes can be used to identify a possible diagnosis for the patient.

The datasets used for the experiment are as follows:

- Asthma vs Bronchitis vs Influenza vs Pneumonia vs Tuberculosis vs URTI
- URTI vs Not URTI
- LRTI vs Not LRTI
- Hypertension vs Not Hypertension

The first dataset tests the capability of the neural network to assess diagnoses with similar symptoms. The next two datasets are for diagnosis predictions against other diagnoses related to it. The final dataset tests the ability to differentiate a diagnosis from a group of interrelated ones.

The classifier used is a Multilayer Feed-Forward Neural Network with Back Propagation. The vectors of tokens containing TF-IDF values are used as input data for the classifier. Datasets for each test case were randomly partitioned such that 75% are included in the training set and 25% for the testing set. The network consists of one input, one hidden, and one output layer. The number of input layer units depends on the number of tokens generated for a given

test case. All datasets are tested on two hidden units schemes. The two schemes for the computation of hidden units are

$$h = \log_2(n) \qquad (1)$$

$$h = \sqrt{n + m} \qquad (2)$$

where *h* is the number of hidden layer units, *n* is the number of input layer units and *m* is the number of output classes. These schemes are patterned from those used by Li et al.[6] to determine the optimal number of hidden units. A supervised back-propagation trainer was initialized using the training data. The trainer parameters are 0.01 learning rate, 1.0 learning rate decay, 0.1 momentum, and 0.01 weight decay. Ten epochs were conducted for the initialized trainer where the percent error of training and testing errors were computed for each epoch. The mean train error and test error for all epochs were computed for each test case to measure the accuracy of the classifier.

4. Results and discussions

The goal of the study is to test the capacity of artificial neural networks for automatic diagnosis prediction using clinical encounter notes. The network was applied to four test cases to analyze different behaviors and to identify a certain best case that will be adopted for implementation. The intelligent agent will be implemented to the consultation and diagnosis workflow of SHINE OS+.

4.1. Similar features on multiple diagnoses

Results show that multiple diagnoses share the same features. This limited the capability of the model to accurately predict among similar classes. For example, 'URTI', 'Pneumonia' and 'Bronchitis' have high occurrences of "cough", "fever", and "runny nose". In this case, the said symptoms are most likely to be associated with the three diagnoses, thus, limiting the network's capability for an accurate prediction. Table 2 shows the total term occurrences of each symptom for each diagnosis.

Table 2. Sample total term occurrences in texts per diagnosis.

Diagnosis	cough	fever	runny nose	colds	yellowish phlegm	dizziness
Asthma	1035	317	379	60	51	29
Bronchitis	5490	2401	2477	112	330	127
Influenza	220	398	233	3	2	35
Pneumonia	2944	1399	1038	211	147	82
Tuberculosis	1475	281	33	133	52	28
URTI	8668	3547	2749	1456	271	277
Total	**19832**	**8343**	**6909**	**1975**	**853**	**578**

Further inspection of the tokenized texts reveals that different diagnoses also share the same combination of symptoms. For example, patients having symptoms of "cough + runny nose" can be associated to having either 'URTI' (389 occurrences), 'Bronchitis' (418 occurrences) or 'Pneumonia' (178 occurrences) due to high number of occurrences. This increases the chance of errors in dis- ease prediction. Table 3 shows the number of times a combination of symptoms appeared for the six interrelated respiratory illnesses.

4.2. Classification using the neural network

Figs. 1 to 4 plot the percent errors over the ten epochs. It can be observed that the fourth test case gives a stable percent error over ten epochs unlike the other test cases which have varying percent errors ranging from 30.00% to above 80.00%. This entails that the network is able to predict 'Hypertension' consistently with minimal errors. For the other test cases, changes in percent error signify that there is uncertainty as to how much the network can predict accurately given a number of iterations. In Table 4, the standard deviations of each test case are presented.

Table 3. Symptoms combination shared by multiple diagnoses.

Diagnosis	cough	cough + runny nose	cough + fever	cough + fever + runny nose
Asthma	108	79	27	26
Bronchitis	456	418	276	315
Influenza	17	29	4	6
Pneumonia	222	178	168	150
Tuberculosis	241	3	17	3
URTI	689	389	266	243
Total	**1733**	**1096**	**758**	**743**

Table 4 shows the average percent error for each test case. It can be noticed that test case 4, i.e. 'Hypertension' vs 'Not Hypertension', garnered the lowest percent errors. It gives percent errors of values only ranging from 10.53% to 11.52%. Testing with the six classes produced the highest percent errors valuing from 54.01% to 65.32%. The six different classes involved contain a very similar set of symptoms as seen in Table 2 and Table 3. Though there is 0.00 standard deviation in the first hidden units scheme, the second hidden units scheme gave the highest standard deviations of 0.1501 for train error and .1525 test error.

In test cases 1 and 3, the model using the first hidden unit scheme was able to predict more accurately having lower percent errors. However, in the second test case the second hidden units scheme presented a more accurate result by only a little margin.

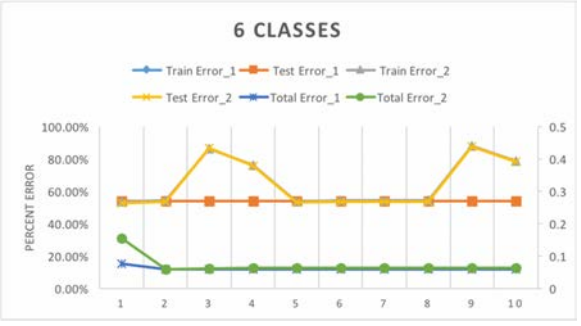

Fig. 1. Train and Test Error Plot for 6-Classes Dataset.

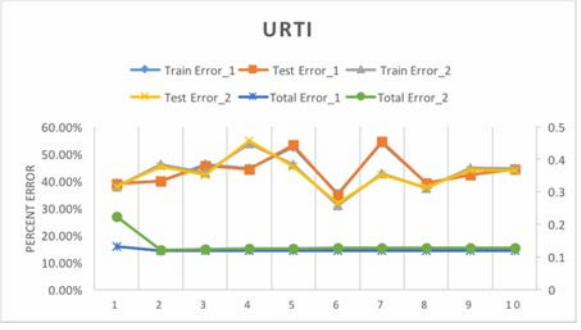

Fig. 2. Train and Test Error Plot for 'URTI' vs 'Not URTI'

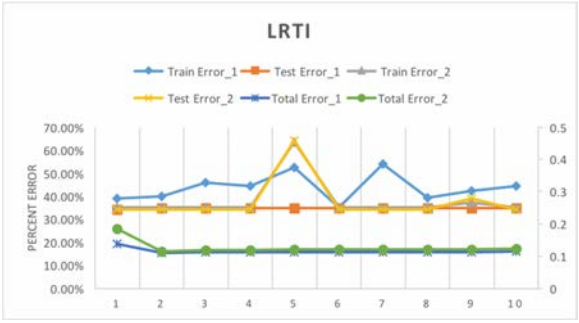

Fig. 3. Train and Test Error Plot for 'LRTI' vs 'Not LRTI'

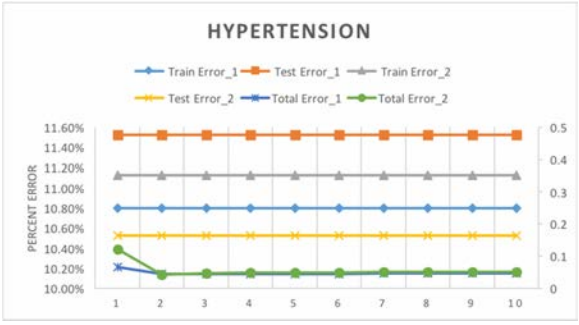

Fig. 4. Train and Test Error Plot for 'Hypertension' vs 'Not Hypertension'

5. Discussion on findings and conclusion

Different behaviors are observed after applying the neural networks on the four data sets. Predicting an unrelated diagnosis versus a group of similar ones yields high accuracy and low percent errors. This signifies that the use of neural networks can already be effective for diagnosis prediction for those having rare signs and symptoms. On the other hand, the percent error becomes higher as the number of interrelated diseases increases, as exhibited in the classification among the six respiratory illnesses. Further developments should be explored for prediction of those with common signs and symptoms. Other specifications of the neural network can also be explored to identify the optimal neural network design.

Table 4. Summary of classification results

Test Case	Hidden Units	Mean		StDev	
		Train Error	Test Error	Train Error	Test Error
1	$\log_2 n$	54.06%	54.01%	0.0000	0.0000
	$\sqrt{n+m}$	65.32%	64.82%	0.1501	0.1525
2	$\log_2 n$	43.74%	43.67%	0.0593	0.0619
	$\sqrt{n+m}$	42.74%	42.49%	0.0620	0.0605
3	$\log_2 n$	35.00%	34.70%	0.0008	0.0015
	$\sqrt{n+m}$	38.21%	37.82%	0.0897	0.0940
4	$\log_2 n$	10.80%	11.52%	0.0000	0.0000
	$\sqrt{n+m}$	11.13%	10.53%	0.0000	0.0000

Initial findings from the experiments show the potentials of applying neural network for automatic disease prediction from clinical text notes. Aside from classification, neural networks can also support other data mining tasks such as feature selection and categorization. The study validates other studies that report the potential of using text

data in predictive modeling. Though there are several intelligent agent systems that are based on purposively designed quantitative analysis, quantifying text data on patient clinical encounter notes shows that this might be an additional feature that can increase the intelligence of a health assistant embedded in EMRs.

6. Recommendations for further studies

It is understood that there will always be nuances in qualitative analysis, especially in the context of doctor's notes, as most symptoms such as fever, cough, and runny nose may lead to similar diagnosis. It is therefore relevant for intelligent agents to also have inputs from standard classification systems such as ICD-10 that can serve as inputs to physicians to further clarify notes. For example, if the notes say: fever, cough, and runny nose, the intelligent agent may ask: for how many days?, how severe?, and other helpful data inputs. It can also be an additional feature for the intelligent agent to have automatic error correction to check for data input spelling, consistency, and other possible areas for errors. Using established clinical notes outside the EMR can also be used to improve on the accuracy of the model, which can then be used as initial knowledge base of the intelligent assistant. Further studies can explore the addition of other data such as patient health records and histories, geographic locations, and possibly hazard data to improve on the relevant feature sets for intelligent agents. Health data consists of various types of data that can be helpful for creating predictive and prescriptive models for health decision support.

Acknowledgements

We would like to acknowledge SHINE OS+, SHINE Labs, Smart Communications, Inc. and the Ateneo Java Wireless Competency Center (AJWCC) for their significant contributions in this study.

References

1. El-Sappagh SH, El-Masri S. A distributed clinical decision support system architecture. Journal of King Saud University-Computer and Information Sciences. 2014 Jan 31;26(1):69-78.
2. Marcelo AB, Cañero JM. Health Information Systems.
3. Simpao AF, Ahumada LM, Gálvez JA, Rehman MA. A review of analytics and clinical informatics in health care. Journal of medical systems. 2014 Apr 1;38(4):1-7.
4. Koh HC, Tan G. Data mining applications in healthcare. Journal of healthcare information management. 2011 Jan;19(2):65.
5. Weiss, Sholom M., et al. *Text mining: predictive methods for analyzing unstructured information*. Springer Science & Business Media, 2010.
6. Gupta V, Lehal GS. A survey of text mining techniques and applications. Journal of emerging technologies in web intelligence. 2009 Jan 8;1(1):60-76.
7. Malakar R. Electronic medical records. Indian Journal of Dermatology. 2006 Apr 1;51(2):140.
8. SHINE OS+. SHINE OS+ about.
9. Byrd RJ, Steinhubl SR, Sun J, Ebadollahi S, Stewart WF. Automatic identification of heart failure diagnostic criteria, using text analysis of clinical notes from electronic health records. International journal of medical informatics. 2014 Dec 31;83(12):983-92.
10. Patra A, Singh D. Neural Network Approach for Text Classification using Relevance Factor as Term Weighing Method. International Journal of Computer Applications. 2013 Jan 1;68(17).
11. Ahmed FE. Artificial neural networks for diagnosis and survival prediction in colon cancer. Molecular cancer. 2005 Aug 6;4(1):29.
12. Li L, Liqing H, Hongru L, Feng Z, Chongxun Z, Pokhrel S, Jie Z. The use of fuzzy backpropagation neural networks for the early diagnosis of hypoxic ischemic encephalopathy in newborns. BioMed Research International. 2011 Jul 24;2011.
13. Zhang Y, Liu S, Zhu Z, Si S. Agent-based intelligent medical diagnosis system for patients. Technology and Health Care. 2015 Jun 17;23(s2).
14. Yılmaz Ö, Erdur RC, Türksever M. SAMS–A Systems Architecture for Developing Intelligent Health Information Systems. Journal of medical systems. 2013 Dec 1;37(6):1-7.
15. Schaul T, Bayer J, Wierstra D, Sun Y, Felder M, Sehnke F, Rückstieß T, Schmidhuber J. PyBrain. The Journal of Machine Learning Research. 2010 Mar 1;11:743-6.