A Bayesian-based Prediction Model for Personalized Medical Health Care

Jiashu Zhao[1], Jimmy Xiangji Huang[2], Xiaohua Hu[3], Joseph Kurian[4], William Melek[5]
[1]*Department of Computer Science and Engineering, York University, Toronto, Canada.*
[2]*School of Information Technology, York University, Toronto, Canada.*
[3]*College of Information Science, Drexel University, Philadelphia, USA.*
[4]*Alpha Global IT, Toronto, Canada.*
[5]*Mechanical and Mechatronics Engineering, University of Waterloo, Waterloo, Canada.*

***Abstract*—In this paper, we present a Bayesian-based Personalized Laboratory Tests prediction (BPLT) model to solve a real world medical problem: how to recommend laboratory tests to a group of patients? Given a patient who has conducted several laboratory tests, BPLT model recommends further laboratory tests that are the most related to this patient. We regard this laboratory test prediction problem as a special classification problem, where a new laboratory test belongs to either a "taken" or "not-taken" class. Our goal is to find the laboratory tests with high probability of "taken" and low probability of "not taken". Based on Bayesian method, the BPLT model builds a weighting function to investigate the correlations among laboratory tests and generate the rank of laboratory tests. In order to evaluate the proposed BPLT model, we further propose a novel evaluation metric to subjectively measure the accuracy of BPLT model. Experimental results show that BPLT model achieves good performance on the real data sets and provides a good solution to our real world application.**

***Keywords*-BPLT, Bayesian Learning, Medical Health Care, Laboratory Test Prediction, Smoothing Technique**

I. INTRODUCTION

With the rapid development of health care systems, a large amount of clinic laboratory test data are available. However, how to make good use of these data is still a challenging problem. In this paper, we aim to solve the problem of predicting laboratory tests for a given group of patients. This research topic comes from a real research project. In reality, it is sometimes hard to track a patient's health condition, but instead general information and laboratory tests are available. Table I is an example dataset with 5 attributes: SDTE (SERVICE DATE), PNUM (PATIENT HEALTH CARD#), PSEX (PATIENT SEX), BDTE (PATIENT DATE OF BIRTH), TSEQ (TEST SEQUENCE NUMBER). For the sake of privacy, the information in Table I is fake and it only shows the format of a class of datasets.

Our goal is to learn from the medial data so that computers can provide recommendations for doctors. We focus on laboratory test recommendation that provides which tests a patient should take. Traditionally, doctors assign laboratory tests for patients based on the symptoms and medical history of patients, as well as the doctors' medical knowledge and experience. Here we are going to propose a laboratory recommendation model, which would subjectively determine whether a laboratory test is related to a patient. The problem that we are going to solve can be described as follows:

SDTE	PNUM	PSEX	BDTE	TSEQ
20120201	patientnumber1	female	01081958	$test_{1,1}$
20120201	patientnumber1	female	01081958	$test_{1,2}$
20120201	patientnumber1	female	01081958	$test_{1,3}$
20120201	patientnumber1	female	01081958	$test_{1,4}$
20120202	patientnumber2	male	11051984	$test_{2,1}$
20120202	patientnumber2	male	11051984	$test_{2,2}$
20120202	patientnumber2	male	11051984	$test_{2,3}$

Table I: An example dataset

Given a set of patients $P = \{p_1, p_2, ..., p_n\}$ and a set of laboratory tests $T = \{t_1, t_2, ...t_M\}$, each patient p_j has done tests $t_{j,1}, ..., t_{j,k_j}$. If a doctor would like to assign a new test for patient p_j, which test in T should be chosen?

Our proposed model analyzes the association among laboratory tests, and therefore estimates the correlation between a patient and a new laboratory test. The intuition behind this model is supported by a medical doctor from Alpha Global IT [1] which provided the medical data. This model could be directly utilized by the doctors in Alpha Global IT.

Computer systems have been used in medical care realm for years [2], [8], [10]. In the first stage, the storage, retrieval, and communication of information are key features of both the practice of medicine and the administration of heath care [7], [11]. Then researchers start to investigate the obtained data. Frick, S. [5] made predictions based on survival and quality of life by nurses and doctors. However, patients are individuals far different from each other, which desires for personalized medical care. Wu, W. H. [12] presented a general architecture for a wearable sensor system that can be customized to an individual patient's needs. Our work in this paper is investigating personalized laboratory tests problem via statistic approaches.

Through analysing the trends of machine learning in medical care, [6] concluded that two directions play important roles: The first one is dealing with reliability of decisions of classifiers. The second one is using machine learning in order to verify some unexplained phenomena from comple-

978-1-4673-2560-8/12/$31.00

mentary medicine. One part of our work belongs to the first topic: How to properly evaluate the reliability of decisions of BPLT? Medical care evaluation systems are various as problems are different. Traditional solution for medical care evaluation is at the level of physician-patient interaction [4], which uses Clinical records for investigating the "Quality of Medical Care". Some newly proposed measurements are based on Statistics, such as [3]. We propose an evaluation metric $CorrectRate_M$ based on MAP [9] in Information Retrieval.

The contribution of this paper is two-fold. First, we make medical predictions based on the laboratory tests, and we introduce Bayesian learning in ranking the laboratory tests. Second, we generate a validation dataset and propose an evaluation metric for models predicting laboratory tests without human interaction.

The remainder of this paper is organized as follows. The basic concept of Bayesian classifier and a Bayesian-based Personalized Laboratory Tests prediction model (BPLT) are introduced in Section 2. We propose an evaluation metric and analyze experimental results in Section 3. Finally, Section 4 presents the conclusion of this paper and some possible future work.

II. Our Proposed Model

In this paper, we make an assumption that the laboratory tests for a patient have correlations with each other. For example, a patient who is suspected to have diabetes, usually take both Glycosylated Hemoglobin (HbA1C) test and Glucose Fasting test. Therefore, there exists relation between Hemoglobin and Glucose Fasting with respect to some hidden information, diabetes in this case. If a patient has already taken Glucose Fasting test, then we recommend to the doctor that maybe Hemoglobin is needed for this patient as well. This is an simple example containing only two laboratory tests. In the rest of this section, we are going to introduce a model that analyzes the correlations among more laboratory tests, and predicts the probability of whether a new laboratory is related to a patient. We regard this test prediction problem as a special classification problem, where a test belongs to either a "taken" or "not-taken" class. We use Bayesian classifier as our basic classifier, and modify it into a personalized ranking model. In the following, we will first introduce the Bayesian classifier.

Please note that in this laboratory tests prediction problem, whether a test should belong to "taken" or "not taken", is different according to various patients. A test is probably taken by some patients but not taken by others, determined by individual patients' physical condition and doctors' judgement. Therefore, a specific test classification model is needed to predict patient's laboratory tests individually. We introduce the idea of a Bayesian-based Personalized Laboratory Tests prediction (BPLT) model in this section, to determine whether a patient should or not take a particular test, given the other tests that he/she has taken.

Denote the events that $test_1, test_2, ...test_n$ are taken by the jth patient as $F_{j,1}, F_{j,2}, ...F_{j,k}$. For example, if we have 5 tests in total, and the a patient who has taken $test_1$, $test_2$ and $test_4$ could be represented as $(F_{1,1}, F_{1,2}, ..., F_{1,5}) = (1, 1, 0, 1, 0)$. The problem is: which new test should be further taken, among all the non-conducted tests? Our work estimates the correlations between new tests and a given individual patient, and then rank the laboratory tests according to their correlations, in order to find the most possible laboratory test that the patient needs.

To define the correlation between a new test $test_0$ and a patient, we utilize the idea of Bayesian Classifier to obtain the probability of F_0, which is the event of taking the new $test_0$, as follows

$$\Pr(F_0|F_{j,1}, F_{j,2}, ...F_{j,k}) \propto \Pr(F_0)\prod_{i=1}^{k}\Pr(F_{j,i}|F_0)$$

On the other hand, the probability of F_0^c, which is the event of not taking a new $test_0$, is

$$\Pr(F_0^c|F_{j,1}, F_{j,2}, ...F_{j,k}) \propto \Pr(F_0^c)\prod_{i=1}^{k}\Pr(F_{j,i}|F_0^c)$$

Since our goal is to find the tests with high probability of "taken" and low probability of "not taken", we define the measurement of the correlation between a patient and a laboratory test based on the probabilities above.

Definition 1: The correlation between a patient p_j and a laboratory test $test_0$ is the log function of the probability of the patient taken $test_0$ divided by the probability of the patient not taken $test_0$.

$$corr(test_0, p_j) = \log\frac{\Pr(F_0|F_{j,1}, F_{j,2}, ...F_{j,k})}{\Pr(F_0^c|F_{j,1}, F_{j,2}, ...F_{j,k})} \quad (1)$$

$corr(test_0, p_j)$ indicates whether p_j should take $test_0$. The higher $corr(test_0, p_j)$ is, the more relevant $test_0$ is to $patient_j$, and with the higher reliability that $patient_j$ should take $test_0$. We can simplify the calculation of $corr(test_0, p_j)$ as follows

$$\begin{aligned} &corr(test_0, p_j) \\ &= \log\frac{\Pr(F_0)}{\Pr(F_0^c)} + \sum_{i=1}^{k}\frac{\log\Pr(F_{j,i}|F_0)}{\Pr(F_{j,i}|F_0^c)} \end{aligned} \quad (2)$$

To future simply the above formula, we make use of the the characteristic of our data. Since a test has only two states "taken" or "not taken", i.e. F_0 or F_0^c, using the property of probability, $\Pr(F_0^c)$ and $\Pr(F_{j,i}|F_0^c)$ in (2) can be eliminated in $corr(test_0, p_j)$, as shown below

$$\log\frac{\Pr(F_0)}{1-\Pr(F_0)} + \sum_{i=1}^{k}\log\frac{\Pr(F_{j,i}|F_0)(1-\Pr(F_0))}{\Pr(F_{j,i}) - \Pr(F_{j,i}|F_0)\Pr(F_0)}$$

From the concept of conditional probability, we know that a joint probability is

$$\Pr(F_{j,i}, F_0) = \Pr(F_{j,i}|F_0)\Pr(F_0)$$

Then we can further simplify

$$\begin{aligned} & corr(test_0, p_j) \\ & = (k-1)\cdot\log\frac{1-\alpha}{\alpha} + \sum_{i=1}^{k}\log\frac{\beta_{j,i}}{\gamma_{j,i}-\beta_{j,i}} \end{aligned} \quad (3)$$

where

$$\alpha = \Pr(F_0) = \frac{number\ of\ patients\ taken\ test_0}{number\ of\ patients}$$

$$\gamma_{j,i} = \Pr(F_{j,i}) = \frac{number\ of\ patients\ that\ F_{j,i}\ holds}{number\ of\ patients}$$

$$\beta_{j,i} = \Pr(F_{j,i}|F_0) = \frac{\Pr(F_{j,i}, F_0)}{\Pr(F_0)}$$

$$= \frac{1}{\alpha}\frac{number\ of\ patients\ that\ both\ F_0\ and\ F_{j,i}\ holds}{number\ of\ patients}$$

We also apply smoothing techniques for the above formula.The next step is ranking $corr(test_0, p_j)$ with respect to all the tests in descending order. The higher a test in the list, the more possible that p_j is related to this test. In other words, the list gives us the recommendation of which test(s) a patient should take.

III. Experimental Results

A. Dataset

The dataset in our experiment is obtained from Alpha Global IT [1]. Alpha Corporate Group is an authorization that has been providing Medical Laboratory, Industrial/Pharmaceutical Laboratory, Diagnostic Imaging services and Managed Care Medical Clinic in addition to providing commercial Electronic Medical Record and Practice Management Software. Interested researchers could contact Alpha Global IT for further information of the datasets. The medical test dataset contains several years patients' records. We utilize 6 months of the records in this paper as a key study, which contain 1,048,575 entries. We investigate the generalized error by randomly splitting the records into a training set and a validation set. We obtain the prior probabilities from the training set, and make predictions on the validation set.

B. Validation Data and Measurement

We propose an approach to obtain a set of validation data by deleting a randomly chosen t^* of each $patient_j$. t^* is stored as golden standard, which will be compared with the predicted laboratory tests to evaluate the system. BPLT model predicts a list of laboratory tests ranked according to (3). To measure the effectiveness of BPLT model, we also propose a new evaluation method for this specific problem, inspired by the idea of Mean Average Precision (MAP) [9] evaluation method in Information Retrieval (IR) domain. BPLT model assigns each laboratory a relevancy weight and ranks all the tests in the database as list $L = t'_{1,j}, ...t'_{k,j}$. By our assumption, a test with the higher relevancy weight will be the more likely to be chosen. We propose the following measurement to evaluate BPLT model.

Training Data Percentage	$CorrectRate_1$	$CorrectRate_3$
60%	0.7074	0.7840
50%	0.6962	0.7837
40%	0.6823	0.7821

Table III: BPLT Performance

Definition 2: The $CorrectRate$ evaluates the accuracy of a BPLT system. It is the number of patients with the desired (golden standard) test matching one of the top M tests generated by the system, divided by the total number of the patients.

$$CorrectRate_M = \frac{\sum_{j=1}^{n} TOP_{j,M}}{n} \quad (4)$$

where

$$TOP_{j,M} = \begin{cases} 1 & if\ \ t^*\ \ matches\ \ a\ \ test\ \ in\ \{t'_{1,j}, ...t'_{M,j}\} \\ 0 & otherwise \end{cases}$$

n is the number of patients, M is a parameter indicating how many top tests are compared to the golden standard test t^*.

Formula (4) gives a rate of correctly predicating t^* in the top M tests of the ranked list L. When $M >= 1$, we regard the prediction correct as long as t^* is within the top M predicted laboratory tests. Since each patient only has one desired test, other measurements such as False Positive Error or Positive False Error can be deducted directly from $CorrectRate_M$. We adopt M = 1 and M = 3, which means the proportion of patients that the golden standard test ranks in the top 1 or the top 3, namely $CorrectRate_1$ and $CorrectRate_3$.

Table II is an example of how the proposed measurement works. Suppose we have 100 laboratory tests and 3 patients in total. A ranked list is obtained from BPLT model for each patient, and compared with the desired test of this patient. " $>$ " is only an symbol indicating that the relevancy weight of the left part is larger than of the right part. We can see that t^* ranks the 1st for $patient_1$, 3rd for $patient_2$, and not within top 3 for $patient_3$, then $TOP_{j,M}$ are shown in the table for given M. The $CorrectRate_1$ and $CorrectRate_3$ have a vale of 0.33 and 0.67 accordingly. Since the patients with t^* ranked in the top 3 will include the patients with t^* ranked in the top 1, $CorrectRate_3$ is always larger or equals to $CorrectRate_1$.

	Desired test t^*	Ranked recommendation tests list	$M = 1$	$M = 3$
$patient_1$	$test_7$	$test_7 > test_8 > test_{40} > ...$	$TOP_{1,1} = 1$	$TOP_{1,3} = 1$
$patient_2$	$test_{18}$	$test_5 > test_3 > test_{18} > ...$	$TOP_{2,1} = 0$	$TOP_{2,3} = 1$
$patient_3$	$test_5$	$test_{63} > test_7 > test_{12} > ...$	$TOP_{3,1} = 0$	$TOP_{3,3} = 0$
All patients	—	—	$CorrectRate_1 = 0.33$	$CorrectRate_3 = 0.67$

Table II: An example of $CorrectRate$

C. Results

The overall performance of our BPLT system is shown in Table III. We conduct experiments on different training sets and validation sets by re-sampling with different training-validation proportions. The percentage of training data we used are 40%, 50%, and 60%. Using 60% data for training has the highest $CorrectRate_1$ and $CorrectRate_3$, which are 0.7074 and 0.7480. Using 50% data for training has the median $CorrectRate_1$ and $CorrectRate_3$, which are 0.6962 and 0.7837. Using 40% data for training has the lowest $CorrectRate_1$ and $CorrectRate_3$, which are 0.6823 and 0.7821. The results show that BPLT could present accurate recommendations of laboratory tests.

IV. Conclusions and Future Work

We propose a BPLT model for personalized medical care, which effectively mines the hidden association among laboratory tests. To evaluate the BPLT model, we propose a new experimental framework and a fair evaluation metric called $CorrectRate_M$. The experiments are conducted on a real laboratory data set obtained from Alpha Global IT. In the experiments, the performance of BPLT is presented, BPLT model can provide accurate recommendations of laboratory tests to the doctors. The experimental results show that we provide an accurate and effective BPLT framework for mining and analyzing laboratory test data.

Using BPLT model, a doctor can find a list of tests where the top ranked tests have more correlations with the patient. The doctor could have a look at the top tests and make further judgement if an extra laboratory test should be added. Therefore, the BPLT model can help the doctors to make a through examinations for the patient, which will finally result to better diagnosing.

In the future, we plan to work on the following issues. The first one is to improve the performance of BPLT and test it on larger data sets which contain more patients and more laboratory tests. The second one is to group the patients according their age and gender, in order to have a through investigation on different patient groups. We are also looking forward to working with medical doctors to analyze the knowledge behind the laboratory tests.

Moreover, another future direction of this work is to analyze the association among the laboratory test results by detecting the co-occurrence of abnormal changes on two or more laboratory tests. The obtained findings will enhance laboratory predication systems.

Acknowledgements

This research is supported in part by the research grant from the Natural Sciences & Engineering Research Council (NSERC) of Canada and the Early Research Award/ Premier's Research Excellence Award. We thank Dr. Joseph Kurian and Dr. William Melek from Alpha Global IT for their help and providing the data.

References

[1] Alpha Global IT. http://www.alpha-it.com/.

[2] D. Bates, M. Cohen, L. Leape, J. Overhage, M. Shabot, and T. Sheridan. Reducing the frequency of errors in medicine using information technology, 2001.

[3] N. Cook. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clinical chemistry*, 54(1):17, 2008.

[4] A. Donabedian. Evaluating the quality of medical care. *Milbank Quarterly*, 83(4):691, 2005.

[5] S. Frick, D. Uehlinger, and R. Zenklusen. Medical futility: Predicting outcome of intensive care unit patients by nurses and doctors-A prospective comparative study*. *Critical care medicine*, 31(2):456, 2003.

[6] I. Kononenko. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*, 23(1):89–109, 2001.

[7] J. Melski, D. Geer, and H. Bleich. Medical information storage and retrieval using preprocessed variables. *Computers and biomedical research, an international journal*, 11(6):613, 1978.

[8] L. Ogiela, R. Tadeusiewicz, and M. Ogiela. Cognitive techniques in medical information systems. *Computers in Biology and Medicine*, 38(4):501–507, 2008.

[9] M. Sanderson. Information retrieval system evaluation: effort, sensitivity, and reliability. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 162–169. ACM New York, NY, USA, 2005.

[10] E. Shortliffe and J. Cimino. *Biomedical informatics: computer applications in health care and biomedicine*. Springer, 2006.

[11] G. Thoma, S. Suthasinekul, F. Walker, J. Cookson, and M. Rashidian. A prototype system for the electronic storage and retrieval of document images. *ACM Transactions on Information Systems (TOIS)*, 3(3):279–291, 1985.

[12] W. Wu, A. Bui, M. Batalin, L. Au, J. Binney, and W. Kaiser. MEDIC: Medical embedded device for individualized care. *Artificial Intelligence in Medicine*, 2008.