Pathway-Finder: An Interactive Recommender System for Supporting Personalized Care Pathways

Rui Liu, Raj Velamur Srinivasan, Kiyana Zolfaghar, Si-Chi Chin, Senjuti Basu Roy, Ankur Teredesai
Center for Data Science, Institute of Technology, UW Tacoma, WA-98402
phone:253-692-5860, fax:253-692-5862
Email: {ruiliu, velamurr, kiyana, scchin, senjutib, ankurt}@uw.edu

***Abstract*—Clinical pathways define the essential component of the complex care process, with the objective to optimize patient outcomes and resource allocation. Clinical pathway analysis has gained increasing attention in order to automate the patient treatment process. In this demonstration paper, we propose `Pathway-Finder`, an interactive recommender system to visually explore and discover clinical pathways. The interactive web-service efficiently collects patient information that is necessary for designing an effective personalized treatment plan. `Pathway-Finder` implements a Bayesian Network to discover causal relationships among different factors. Additionally, the system implements a big-data infrastructure using Spark that is hosted as a HDinsight cluster on Microsoft Azure for Research platform to support real-time recommendation and visualization. We demonstrate `Pathway-Finder` to interactively recommend personalized interventions to minimize 30-day readmission risk for Heart Failure (HF).**

I. Introduction

Clinical pathways are widely used by hospitals for automating and managing patient treatment process. Effective clinical pathway analysis captures best clinical practices that contribute to satisfactory outcome, such as length of stay of each patient [1], [2]. The development of clinical pathways is a time-consuming process, requiring the collaborations among physicians, nurses and staffs in a hospital. On the other hand, the execution of clinical pathways needs to be adaptive to support clinician's information needs and to accommodate individual differences of each patient. Usually, when a new patient is admitted to a hospital, little information is known about the patient and numerous questions can be asked of the patient. In order to use the limited amount of visit time wisely[1], clinicians need to acquire necessary information about the patient in a timely manner to inform treatment plans and develop clinical pathways.

We develop `Pathway-Finder`, a novel interactive recommender system for clinical pathways analysis, to identify and gather necessary information for supporting clinical decision making and to recommend appropriate interventions that can lead to improved care quality. The proposed interactive system enables the user to enter factors and visualize the connections between patient demographic characteristics, disease conditions(comorbidities or diagnoses), possible interventions, and the targeted clinical outcome. `Pathway-Finder` is an information system and clinical decision support system that helps clinicians to indicate which relevant information is needed to provide quality care experience. `Pathway-Finder` informs clinicians what are the necessary information to gather in order to improve the outcome of the care. In this demonstration, we use a heart failure cohort provided in [4] to minimize the 30-day risk of readmission for Heart Failure (the targeted outcome). However, the system is flexible enough to adjust to any user specified clinical condition and the outcome of interest.

The contributions of `Pathway-Finder` are as follows:

- First, the system provides interactive discovery and exploration of clinical pathways analysis;
- Second, the system iteratively collects necessary patient information that drive the development of treatment plan;
- Third, the system visualizes the trace and predicted outcome of a patient, supporting personalized intervention recommendation;
- Last, the system implements a big-data infrastructure using Spark that is hosted as a HDinsight cluster on Microsoft Azure for Research platform, supporting speedy real-time interactive visualization.

The modeling and the solution relies on learning the structure and the probability distribution of a *Bayesian Network* [5] from the available patient data. As the number of attributes (or attribute values) increases, the Bayesian Network grows in size, resulting in exponential number of look ups to perform in order to recommend interventions. To recommend interventions real time, the system, therefore, makes use of a distributed solution by developing a scalable cloud-based infrastructure hosted on Windows Azure for Research.

Section II describes the technical specifications of our proposed system. `Pathway-Finder` uses Bayesian Network learning for offline computation (Section II-A) and provide a scalable key-value structure to store exponential number of conditional probabilities to support online real-time factor retrieval for the visualization (Section II-B). Section III demonstrates the four stages of `Pathway-Finder` with a basic use case. We conclude the paper in Section V.

II. Technical Specifications: System Overview

`Pathway-Finder` is a cloud based web-service hosted on Microsoft Azure for Research platform. The objective is to interactively discover more about the user health conditions

[1]The average visit duration with physicians less than 20 minutes [3].

and adaptively recommend care-pathways to minimize her 30-day readmission risk for heart failure. The majority of the proposed system components in this demonstration are pre-computed and stored to increase the speed of the application. Figure-1 provides the overview of the system that comprises of offline and online layer. The UI enlists simple socio-demographic factors and the user selects respective values for those from the drop-down. After that, the system alternatively suggests a set of diagnoses and interventions (utilities and procedures) to the user and then she/he selects some of them. In Section II-A we describe the offline computations, and Section II-B is used to describe the computations that take place, once the user start interacting with the system.

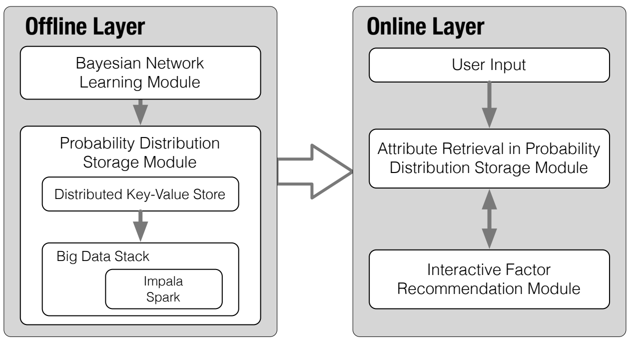


Fig. 1: Architectural Overview of `Pathway-Finder`

A. *Offline Computation*

The offline computation is performed using two modules. (a) Bayesian Network Learning Module, and (b) Key Value Storage Module. We describe these two modules next.

(a) Bayesian Network Learning Module: At the heart of the system `Pathway-Finder`, there exists an intervention recommendation module which is mostly designed offline. The objective of this module is "discover" the causal relationship between different factors (or attributes) related to heart failure readmission and how their interplay impacts the readmission risk. We briefly describe our solutions next.

A hierarchical Bayesian network effectively depicts the causal relationships between the factors and how their interplay relates to lowering the heart failure readmission risk. Thus, we model the intervention recommendation as a network learning task using the Bayesian network learning principles. For structure learning, our designed solution appropriately adapts Constraint Based Bayesian Network Learning algorithm [6], [7], *Score Based Learning Algorithm* [8] and *Hybrid Algorithm* [9] that combines both Constrained based and Score based approaches. Once the structure is defined, we use parameter learning [10], [11] techniques to learn the probability distribution at each node. In our implementation, we use Bayesian Parameter Estimation [12] to learn the parameter θ. The created probability distributions will go through the Probability Storage Module after that.

(b) Probability Distribution Storage Module: In this section, we present methods to transform the exponential number of conditional probabilities learned from Bayesian Networks (Section II-A) to create a scalable storage module for efficient retrieval.

For the simplicity of exposition, imagine that a Bayesian Network has learned that "Gender" (M/F) and "Race" (imagine "Caucasian" and "Pacific Islanders" are two possible race values) has *causal relationship* with "Heart Failure (HF)". The task is to store the probability distribution function (Pdf in short) of each of these three nodes in a *distributed key-value store*. The keys are multiple set of composite keys consisting of the minimum number of combination of the attributes required to maintain the uniqueness and have a cascading relationship with each other. The value are the probability of getting diagnoses, given key and the rest of the attributes. For our example above, the first set of keys are the two "Gender" values with rest of the attribute combinations as the value (such as, $Pr(HF|M, Caucasian), Pr(HF|M, PacificIslander)$ (similarly for females). The second set of keys will be for different ("Gender", "Race") combinations and the values are the probabilities of HF given the keys. The distributed key-value store is built on top of a big data stack like Spark [13]. This stack is hosted as a HDinsight cluster on Azure.

B. *Online Computation*

The online layer is designed with three modules that are described next.

(a) User Input: The interface accepts simple socio-demographic attribute values from the user as it is shown in stage-1 of Figure 3a. After that, the two other modules, iteratively interact with the user.

(b) Probability Distribution Lookup Module: This module is invoked multiple times to do look up either for the diagnoses or for the interventions. Based on the user input, the search goes inside the Probability Distribution Lookup Module to retrieve either the diagnoses or the interventions that the user is most likely to have.

The look up from the distributed key-value store checks if the key is present in the store or not. Based on the user input, various combinations of keys are formed. Once the conditional probability for all possible diagnoses as entered by user have been looked up, the intervention will be recommended for the diagnosis with the highest probability. For example, if a user enters "Gender=M",and "Race = Caucasian", then the first look up is on "Gender". The second look up is on "Gender" and "Race", both. The lookup continues until the conditional probability for the diagnoses are retrieved as shown in the Figure-2. If the key is not present in the store, then based on the user's input the most similar key-value pair will be retrieved.

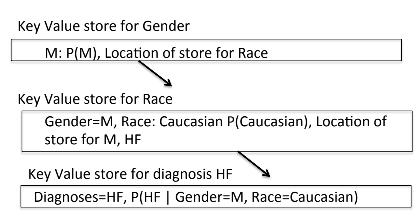


Fig. 2: Distributed Key-Value Lookup on Spark

For the above example, imagine there are 5 diagnoses/comorbidities (DX1,DX2,DX3,DX4,DX5) that the user may have and `Pathway-Finder` wants to suggest the top-3 likely ones to the user. For this, the call comes to this module to compute the following 5 probabilities.

$$Pr(DX1|Gender = "M" \; and \, Race = "Caucasian")$$
$$Pr(DX2|Gender = "M" \; and \, Race = "Caucasian")$$
$$Pr(DX3|Gender = "M" \; and \, Race = "Caucasian")$$
$$Pr(DX4|Gender = "M" \; and \, Race = "Caucasian")$$
$$Pr(DX5|Gender = "M" \; and \, Race = "Caucasian")$$

The module ranks all the diagnoses/comorbidities based on individual probabilities and suggests top 3 highest diagnoses as most likely to investigate as described next.

(c) Interactive Exploration/Recommendation module: There is an iterative interaction between this module and the lookup module. Based on the user selected values so far, the system iteratively suggests more interventions or discovers more diagnoses for her, until one of the stopping conditions is reached. For the example scenario, out of three diagnoses (imagine those are DX2, DX4, DX5) that are suggested to the user, she only selects DX4. Then, based on that the most appropriate set of interventions (let us say top-3) is selected to her (by making a call to the look up module) to retrieve the interventions that minimizes her readmission risk the most.

III. SYSTEM DEMONSTRATION

In this section we demonstrate the four stages of `Pathway-Finder`: 1) Initial input collection; 2) Comorbidities exploration; 3) Intervention recommendation; 4) Outcome prediction and intervention adjustment loop. Figure 3 presents the progress of each stage. We believe that the system is useful to the patients, as well as to the clinicians and physicians.

The system is showcased using State Inpatient Databases (SID)[2] of Washington State of year 2010 and 2011. The HF cohort contains 3,908 distinct diagnosis codes and 2,049 procedure codes and total of 119,988 patient records.

A. Initial Input Collection

At the first stage, we collect preliminary information from user's input in an interactive way. We collect the patients' basic information first. `Pathway-Finder` proceeds with the limited information provided by the user, displaying associated comorbidities from the learned Bayesian Networks described in Section II. For example, as shown in Figure 3a, a user provides her age(=70-74), gender(=Female), ethnic group(=African American) which are orange circles, leaving the others blank. Provided with the preliminary inputs, the network is expanded with a set of corresponding diagnosis/comorbidities, which are Anemia, Congestive Heart Failure, Renal Failure, and Hypertension. The thickness of the links between factor nodes implies the probabilities of one factor leading to the other. The thicker they are, the stronger the connections. The user can interact with the form or directly interact with the network by clicking on the nodes. After clicking the "Submit" button, we enter the stage 2 of the system.

[2]http://www.hcup-us.ahrq.gov/sidoverview.jsp

B. Comorbidities Recommendations

At the second stage, the user can add diagnosis/comorbidities information. For example, as shown in Figure 3b, she clicks on the two circles that turn green: Congestive Heart Failure and Renal Failure. Once the user click "Submit" at this page, our system will show her the appropriate interventions based on all the information collected and other diagnosis or comorbidities that could also appear for this specific patient. In this case, the interventions we recommended are Blood Processing, Echocardiograms, CT Scan, Cardiac Stress Test, Emergency Room, and Physical Therapy (gray circles in Figure 3b). Meanwhile we encourage more inputs from users by suggesting that the patient could also have Chronic Obstructive Pulmonary Disease (COPD) and Fluid and Electrotype (Lytes) Disorder. The additional suggestions were learned from our data and can be used to improve the prediction accuracy of our system. This stage is to assist clinicians to identify the appropriate interventions and other diseases the patient of interest might have.

C. Intervention Recommendations

At the third stage, a clinician can continue filling information about the patient and select appropriate interventions. In our case, the clinician selects Echocardiograms, Cardiac Stress Test, and Physical Therapy (the green circles in the third layer in Figure 3c). On the other hand, she finds the patient actually have Fluid and Electrotype (Lytes) Disorder, so she add this new diagnosis the the network, which is the green circle in the second layer. Based on all the user inputs in from the first three stages, our system predicts the Readmission Risk of the patient. As shown in Figure 3c, `Pathway-Finder` estimates 30% probability that the patient of interest will be readmitted.

D. Intervention Adjustment

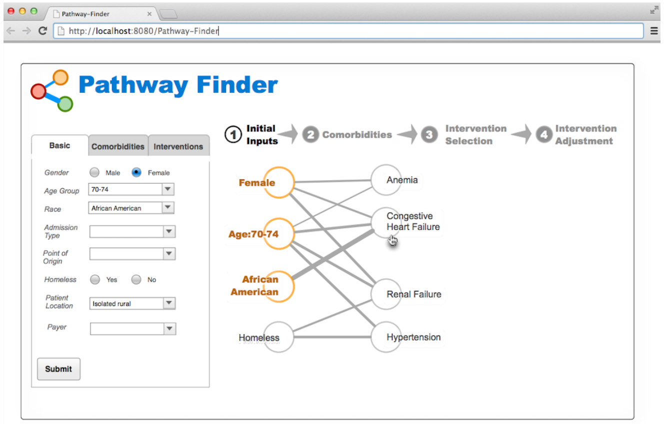
At the last stage, `Pathway-Finder` allows clinicians to adjust their intervention strategies to preview the impact of the new treatment plan to the targeted outcome. In our example shown in Figure 3d, the clinician adds CT Scan and observed a reduced readmission risks to 17%. This stage provides an iterative process to support a loop of discovery between interventions and the targeted outcome.

IV. DEMONSTRATION LOGISTICS

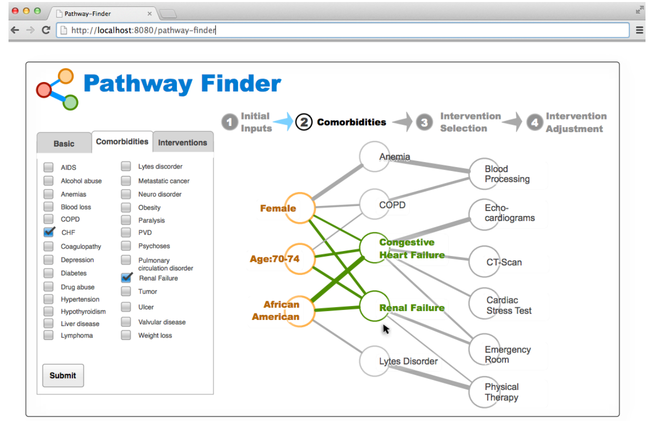
The URL for `Pathway-Finder` will be available before the beginning of ICDM conference. For the purpose of demonstration at ICDM conference, we will provide a laptop to access the system. We will require access to internet, a table that is large enough to accommodate laptop and poster board, and convenient access to power outlets for our laptop. Currently, we are finalizing the implementation of the offline layer of `Pathway-Finder` detailed in Section II-A. We are in the process of implementing the big data store for the online layer (as detailed in Section II-B). We expect the final completion of the UI and the entire recommender system by September 2014.

V. CONCLUSION

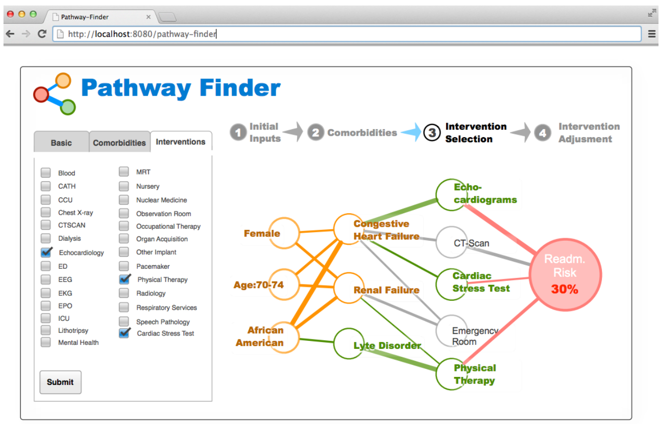
We propose an *interactive system* called `Pathway-Finder`, with the objective to visually explore, discover, and recommend clinical pathways for



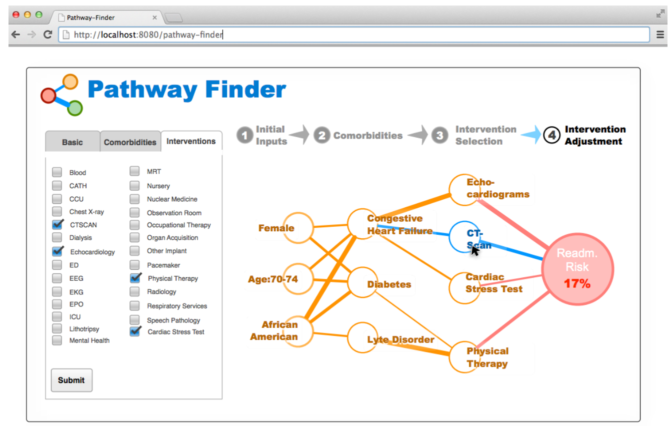
(a) Stage 1: Initial Input Collection



(b) Stage 2: Comorbidities Recommendation



(c) Stage 3: Intervention Recommendation



(d) Stage 4: Intervention Adjustment

Fig. 3: Screen Shots of `Pathway-Finder`

health conditions. We demonstrate `Pathway-Finder` to interactively recommend interventions to minimize the readmission risk due to Heart Failure (HF). At the heart of `Pathway-Finder`, there exists a Bayesian Network that learns the causal relationship among different factors and how that contributes to HF readmission risk. Further novelty of the system includes a distributed implementation using Spark hosted on Microsoft Azure for Research Platform, which is capable to perform real time lookup over the learned network to interactively recommend interventions. Our demonstration also involves a high dimensional real patient dataset with hundreds and thousands of records and several hundreds of factors. To the best of our knowledge, `Pathway-Finder` is the first system that is empowered with the ability to interactively recommend and explore pathways for different clinical conditions.

References

[1] Z. Huang, W. Dong, L. Ji, C. Gan, X. Lu, and H. Duan, "Discovery of clinical pathway patterns from event logs using probabilistic topic models," *Journal of Biomedical Informatics*, vol. 47, pp. 39–57, Feb. 2014.

[2] H. Iwata, S. Hirano, and S. Tsumoto, "Construction of clinical pathway based on similarity-based mining in hospital information system," *Procedia Computer Science*, vol. 31, pp. 1107–1115, 2014.

[3] A. Gottschalk and S. A. Flocke, "Time spent in face-to-face patient care and work outside the examination room," *Annals of Family Medicine*, vol. 3, no. 6, pp. 488–493, Nov. 2005.

[4] R. Liu *et al.*, "A framework to recommend interventions for 30-day heart failure readmission risk," *CoRR*, 2014.

[5] J. Han and M. Kamber, *Data mining: concepts and techniques*, 2006.

[6] J. Pearl and T. S. Verma, "A theory of inferred causation," *Studies in Logic and the Foundations of Mathematics*, 1995.

[7] J. Cheng and R. Greiner, "Comparing bayesian network classifiers," in *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1999, pp. 101–108.

[8] F. V. Jensen and T. D. Nielsen, *Bayesian networks and decision graphs*. Springer, 2007.

[9] I. Tsamardinos, L. E. Brown, and C. F. Aliferis, "The max-min hill-climbing bayesian network structure learning algorithm," *Machine learning*, vol. 65, no. 1, pp. 31–78, 2006.

[10] D. Heckerman, D. Geiger, and D. M. Chickering, "Learning bayesian networks: The combination of knowledge and statistical data," *Machine learning*, vol. 20, no. 3, pp. 197–243, 1995.

[11] G. E. Box and G. C. Tiao, *Bayesian inference in statistical analysis*. John Wiley & Sons, 2011, vol. 40.

[12] S. C. Kramer and H. W. Sorenson, "Bayesian parameter estimation," *Automatic Control, IEEE Transactions on*, 1988.

[13] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, "Spark: Cluster computing with working sets," ser. HotCloud'10.