A Hybrid Health Journey Recommender System using Electronic Medical Records

Soheil Jamshidi, Ali Torkamani, Jynelle Mellen,
Malhar Jhaveri, Penny Pan, James Chung, Hakan Kardes
Cambia Health Solutions
Portland, OR, 97201
{firstName}.{lastName}@cambiahealth.com

ABSTRACT

We present a recommender system aimed at improving the healthcare experience of consumers. Our model provides actionable insights to cohorts or individuals, based on their collective and personal health-related data. The actionable insights are delivered through digital interventions to help prevent adverse events for the consumer. By proposing timely and personalized suggestions, we will improve consumer health outcomes and prevent complications, which would also result in cost-savings. Our recommendation system employs an ensembling technique, where at its core, we have a Bayesian network that uses administrative claims data but could be extended to use Electronic Health Records (EHR) data for learning the structure of the interwoven health graph (conditions, medications, procedures, and more). This method allows for predicting the probability of various outcomes conditioned on the consumers' evidential health data. We also couple our ensemble method with a shallow random forest model to further refine the personalized recommendations after receiving the consumer's feedback. The experimental results show that our system significantly improves the precision-recall metrics of several intervention targets compared to a random baseline.

CCS CONCEPTS

• Applied computing → Consumer health; **Health care information systems**; **• Information systems → Recommender systems**;

KEYWORDS

Health Journey; Health Recommender System; Hybrid Recommender System

ACM Reference Format:
Soheil Jamshidi, Ali Torkamani, Jynelle Mellen,, Malhar Jhaveri, Penny Pan, James Chung, Hakan Kardes . 2018. A Hybrid Health Journey Recommender System, using Electronic Medical Records. In *Proceedings of the Third International Workshop on Health Recommender Systems co-located with Twelfth ACM Conference on Recommender Systems (HealthRecSys'18), Vancouver, BC, Canada, October 6, 2018* , 6 pages.

1 INTRODUCTION

Availability of clinical data in form of Electronic Health Records (EHR) has dramatically increased over the past decade. In addition to EHRs, there are large volumes of administrative payment data that offer insights to researchers. However, administrative data and EHRs are scattered among numerous entities and sources, such as health plans, laboratories, providers, hospitals, chart notes, and more. In addition to the disparate sources, the breadth, depth, linkage, and scale of the data lead to further complexity. For end-users (consumers) it can make interpretation difficult due to information overload or a lack of information, as well as term inconsistency. The increasing need to leverage the health records led to the presence of Health Recommender Systems (HRS) [4, 13]. Such recommender systems can target medical experts or patients and play a vital role in improving an individual's health by providing insightful recommendations. These systems are primarily created to handle ambiguous diagnosis situation because of the varied decisions of providers [13] for certain diseases. In that case, recommenders created by medical specialists provide insight for tailored diagnosis procedure for patients. To this aim, Machine Learning (ML) methods act as enablers. There has been a large number of studies on a wide range of machine learning techniques - such as decision trees, multi-layer perceptron (MLP), support vector machine (SVM) - that has been focused on a variety of diseases - dementia, kidney, and heart diseases to name a few [1–3, 10, 14–16, 18].

In this paper, we present a consumer-focused recommender system that will give individuals suggestions based on their collective health-related data (EHR and claims). To accomplish this, we leverage an ensemble algorithm, where a Bayesian network (BN) is combined with a random forest (RF). Probabilistic Graphical Models such as BN are known to tolerate the data uncertainty (noise, ambiguity, and missing values) in a consistent and mathematically correct way [25] in inference phase and RF facilitates refining the personalized recommendations after receiving the consumer's feedback. Such a system learns the conditional probability table using maximum entropy or belief propagation approach on large datasets derived from over a million records with thousands of diagnoses and findings, and over a hundred variables. Medical and pharmacy claims in combination with laboratory results, nurse notes, and consumer data form a substantially powerful aggregated dataset that can be used to train a model that offers actionable recommendations. This recommendation system can serve a variety of health-related applications (cost predictions and engagement modeling for example) or can be presented as a product to the market.

The structure of the remaining paper is as follows. Section 2 introduces the features of the data. Section 3 demonstrates the methods used for learning and inference. Section 4 explains the technical details of our graphical model. We discuss the experimental results in Section 5. Then, in Section 6 we review the existing techniques

HealthRecSys'18, October 6, 2018, Vancouver, BC, Canada

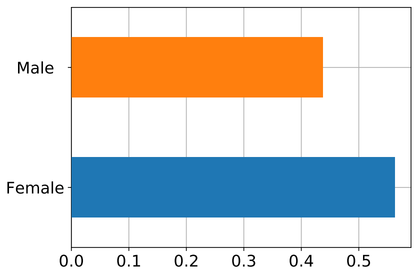

Figure 1: Gender Distribution

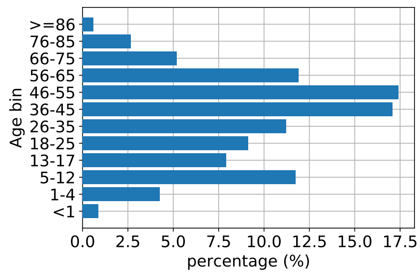

Figure 2: Age Group Distribution

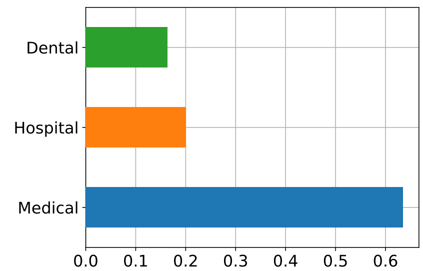

Figure 3: Claim Types Distribution

and studies. We conclude the paper with a discussion and an outline of the directions for further research in Section 7.

2 NATURE OF DATA

The dataset used in our recommender system is largely derived from the claim data with over 1.5 million claims and over a hundred distinct attributes over a 6 year period.

Available attributes can be grouped as numerical and categorical attributes. For example, consumer *gender* has two states, *male* and *female*. Figure 1 illustrates that more than 55% of our target group are women. We divided *Age* into twelve bins as follows: ≤1, 1-4, 5-12, 13-17, 18-25, from 26 to 85 in 10-year bins, and ≥86. It was confirmed by domain experts that patients in these age ranges generally develop similar conditions. The distribution of the number of consumers per age bin is shown in Figure 2. The claim records cover 4 types of data sources: dental, medical, hospital. As it is shown in Figure 3, 65% of the claims are related to medical claims. We term the observed data, the manifests. The manifests are computed as aggregations based on meaningful categories. For each manifest, there is an integer count ≥ 0 that signifies the number of times a person had the event in a quarter. Currently, our system only relies on 4 major categories; drug classification, diagnosis classification, providers specialty, and service category. Finally, since each patient is associated with many records (claims) we aggregated records of each patient by calendar quarters (as suggested by domain experts to be the most relevant time frame to capture related health events). However, our system can be used for different time granularities.

3 METHODOLOGY

We rely on a hybrid approach leveraging Probabilistic Graphical Models (PGM), Random Forest (RF), and Collaborative Filtering (CF) technique to obtain a vector of recommendations and combine the results using an ensembler. This way, we benefit from the power of PGMs in capturing the propagation of effects, CF in considering the similar situations, and RFs targeting tailored recommendations. Our proposed framework is illustrated in Figure 4 where data from different sources is fed into analysis block where we train and use our models. The output is a list of recommendations delivered via a mobile application. Based on the feedback we gather from the users (consumers, providers, etc.) on the quality of the recommendations, we can optimize the weights of the ensembler. In this study, we focus only on the green boxes, PGM and RF and the rest will be touched in our future works.

PGMs can handle large datasets in a computationally tractable manner. We leverage a *Bayesian network* on a Directed Acyclic Graph (DAG). Before learning the conditional probabilities of the network, we found it useful to transform the data in a way that each patient observation corresponded to the manifests exhibited in a quarter, which is long enough to cover a sequence of symptoms, that let us focus on a less noisy sequence of healthcare events. We consider populating the data for the immediate previous quarter and the next quarter for every quarter in our dataset, and hence there are triple of the observations (or rows) for any given patient.

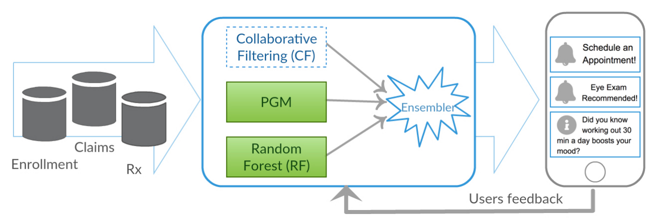

Figure 4: Proposed framework for our recommender system

3.1 Structure learning

Given the transformed data, we can translate the questions to the following prediction/inference problem: Let a patient with a total of N features (manifests) including his/her medical tests, drugs, health events, and other manifests related to a period, the previous period and the next period. If we observe x features out of these N features, can we predict the values (or the probability distributions) of the remaining $N - x$ features based on the available historical data. Using this data as input we learn the structure and create the model. The steps involved are as follows:

First, to derive a joint probability distribution table, we transformed the input matrix to a discrete form with $0, 1$ states. The input data had the number of times a manifest was observed for a patient in a given quarter. If this value was non-zero, we replaced it with a 1. The matrix now represents if the manifest occurred at least once in the period.

Then, we convert the matrix to have manifests observed in a quarter along with manifests observed in the next and previous quarters. This transformation provided us with the data structure with which we could predict or infer manifests of a quarter given those from another quarter.

We consider the data as a matrix A. Then, we find $A^T \times A$ where A^T represents the transpose of A. The resulting symmetric matrix B has the number of joint occurrences of manifests across all patients. The diagonals in this matrix represent the number of occurrences of the manifest for all of the patients. Dividing the row values by the diagonal value in the row resulted in the conditional probability of the column manifest, given the row manifest, $\frac{P(MR,MC)}{P(MR)}$ where MR is the manifest in the row, MC is the manifest in the column, resulting in $P(MC|MR)$. As expected, all the diagonals reduce to 1, and the resulting matrix is no more symmetric.

To determine the significant relationships and to discard those that were not as significant, we set a threshold of 0.05 (5%) for the conditional probability. If a conditional probability was greater than 5% we retain it. This choice of threshold was arbitrary, and as an improvement, we should consult with our domain experts to verify the structure and adjust accordingly. Each relationship with a conditional probability above the threshold represents a directed edge in a graph, with the arrow going from the row manifest to the column manifest. Then, we remove cyclic relationships (the diagonal entries because Bayesian model does not allow loops) using the *networkX* python library. This function detects cycles on a first-come basis and removes the last encountered edge once a cycle is detected. These edges form a DAG structure.

The fit function estimated the *Conditional Probability Distribution* (CPD) for each variable based on the given data and the parameter estimation approach we use. In our case, we use the Bayesian parameter estimation because it considers the probability distribution representing our prior knowledge (how likely are we to believe in the different choices of parameters) and the support of the data (because confidence increases with more data). Moreover, our prior distribution is not uniform and hence this is also a reason to use Bayesian parameter estimation. Since our aim is to predict the values of an unknown manifest we fit the model with the training data set. At this step, the given data in graphical form was ready for performing various types of reasoning.

3.2 Hybrid Scoring

As shown in Figure 4, we use multiple models to obtain the final recommendation. One approach is to apply weighted ensemble methods to obtain better predictive performance than could be obtained from each of these models independently. The final recommendations can be used for a wide range of applications. Here, we focus on a Mobile App that provides health-care *benefit* or *educational* recommendations. For example, if a high probability of hypertension is predicted, then the App would recommend that the person visits his/her Primary Care Physician (PCP).

3.3 Feedback loop

Feedback is a valuable asset for personalizing the recommendations as well as making better recommendations to similar people. The feedback loop can directly contribute to updating the weight vector of the ensemble method, as well as hyperparameter tuning of the individual models. In fact, we need an online learning algorithm to incorporate the feedback into the re-training phase. However, not every feedback or data observation has the same weight/quality. Therefore, we need to consider a context-aware algorithm to pay more attention to the more important segments of information. Attention modeling will be done as a part of our future work.

4 TECHNICAL DETAILS

In this section, we briefly review the technical aspects of our Bayesian network and how parameter learning methods are used to estimate the conditional probabilities for the given set of claims and predict the state or occurrence of a new set of claims.

4.1 Bayesian Network

Bayesian network (BN) is a probabilistic graphical model (PGM) that represents a set of random variable (nodes) say $X_1, \ldots.., X_n$ and their conditional dependencies (edges corresponding to direct influence of one node on another), say $X_1 \perp\!\!\!\perp X_3|X6$, using a directed acyclic graph (DAG). By surfacing these independencies we can reduce the number of values needed to be stored in order to represent the joint probability distribution and thus makes the representation more compact.

For our purpose, we use two layers of inference: *structure* and *parameter learning*. By leveraging structure inference, we create the skeleton using conditional probabilities and domain expert input which captures the dependencies between the variables. The second layer utilizes the dependencies and historical data to estimate the conditional probability distributions of the individual variables.

In parameter learning, there are two main methods:

- Maximum likelihood estimation
- Bayesian estimation

We use Bayesian estimation over maximum likelihood estimation (MLE) because MLE considers a uniform prior distribution and this might lead us to end up in wrong conclusions about the likelihood of a variable θ_{X_i} and adjust the likelihood based on whether the sample is biased or not. Also, MLE does not update the confidence of θ_{X_i} with the change in the size of the data (450,000 out of 1,000,000 vs 45 out of 100). Thus, in Bayesian estimation, we use the prior knowledge about θ with its probability distribution. This distribution will represent how likely we believe the different choices of parameters. Therefore, we can create a joint distribution, which captures the assumption over the parameters θ and the data we are to observe. Each new data point gives us more information about θ and hence the probability of the next occurrence. Hence the posterior distribution for Bayesian estimation is:

$$\Pr(\theta|x[1], .., x[M]) = \frac{\Pr(x[1], .., x[M]|\theta)\Pr(\theta)}{\Pr(x[1], .., x[M])} \tag{1}$$

4.2 Inference

Finding the conditional probability distribution (CPD) over some variables $\Pr(Y|E = e)$ is the same as inferring from a model. Therefore, predicting values for a new data point is the same as finding the conditional probability of the unknown variables, given the observed values of other variables. The CPDs can be computed from the joint probability distribution of the variables, by marginalizing and reducing them over variables and states.

In addition, we are interested in finding the state of a set of variables given other set of variables. It is simply an inference query over the model and state having higher probability would be the prediction by the model. However, computing the joint probability distribution will give us an exponentially large table, which the probabilistic graphical model helps to avoid these tables. There are two algorithms we can use for inference:

- Variable elimination
- Belief propagation

We use variable elimination over belief propagation since the former is suitable for a very large network as it is not memory-expensive. In addition, we discard the generated intermediate factors and hence it is more flexible than BP. In variable elimination, consider the model $A \to B \to C \to D$ and we try to find $\Pr(D)$:

$$\Pr(D) = \sum \Pr(a) \Pr(b|a) \Pr(c|b) \Pr(D|c) \tag{2}$$

In variable elimination, we can sum over parts of the product instead of over the complete product. Hence, Eq. (2) becomes:

$$\begin{aligned} \Pr(D) &= \sum_a \sum_b \sum_c \Pr(a) \Pr(b|a) \Pr(c|b) \Pr(D|c) = \\ &\sum_c \Pr(D|c) \sum_b \Pr(c|b) \sum_a \Pr(a) \Pr(b|a) \end{aligned} \tag{3}$$

This method helps to significantly reduce the computation required to compute the probabilities. Hence variable elimination is much more efficient for calculating probability distributions than normalizing and marginalizing the joint probability distribution.

Our prediction function uses a maximum a posteriori probability to find the states of variables corresponding to the maximum probability in the joint distribution. This is useful when we want to predict the state of variables in our model. Moreover, we introduce another operation on factors called maximization. Maximum a posteriori query is essentially a way to predict the state of variables, given the state of other variables. Thus, using the trained model, we try to predict the states of variables for new data points. To design the models, we need to create conditional probability distributions or factors, add them to the base model, create an inference object, and then do maximum a posteriori queries over it for new data points to predict variable states.

5 EXPERIMENTAL RESULTS

While our un-targeted Bayesian network based recommender system can be leveraged to address a wide range of questions, in some cases a targeted model (such as a random forest) can be more beneficial. In our proposed framework depicted in Figure 4, we have included both components. To assess the abilities of our recommender system, we focus on recommendations of *vision* and *dental* benefits and train the *random forest* model on the same dataset and set of features that we trained our probabilistic graphical model and discuss the results in this section.

We trained the RF model using the default parameteres except for the followings: *n_estimators*=30,*max_depth*=350, *random_state*=0, and *min_samples_leaf*=2. Before fitting the data to the model we split the data into random train and test set in 80:20 ratio (widely recommended split ratio). The test set does not contain the columns/-manifests that we are interested in predicting. The shape of the data used was 2,010 cases with around 1.5K features. We used a subset of the data set because of the constraint in the computational power. Figure 5 and Figure 6 illustrate the Precision-Recall ROC for dental and vision benefits, respectively. In each figure, the expected result from a random guess - based on the frequency of the positive class that is 27.38% (4.66%) for dental (vision) benefit - is also depicted.

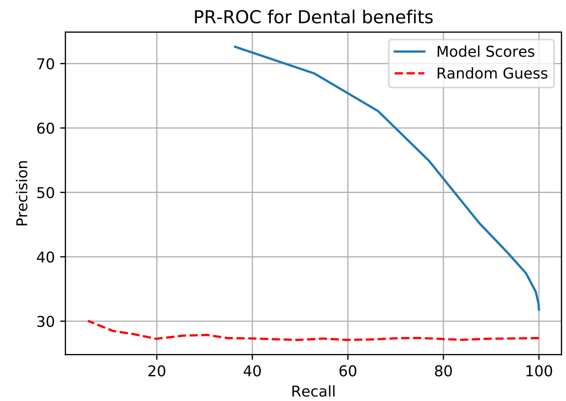

Figure 5: ROC measure of Random Forest on Dental benefits

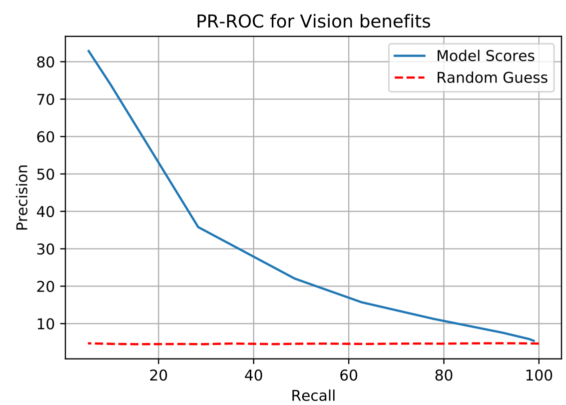

Figure 6: ROC measure of Random Forest on Vision benefits

Using the trained Bayesian network, we predicted the states of all the missing columns/features of the test set. Prediction is done by belief propagation where we find the most probable state of the unknown manifests/features given the states of the other manifest-s/features using CPD's. Fig 7 shows the correlation of manifests related to a specific target (Diabetes mellitus without complication, in the next quarter) and their pairwise correlations considering two consecutive quarters. As expected, having Diabetes mellitus without complication and taking *antidiabetics* in the previous quarter have a strong correlation with having it in the next quarter. Interestingly, the correlation between hypertension, the disorder of lipid, and high Glucose in blood are also captured by our model.

The probabilities gathered from our Bayesian network can be represented as a network. In Fig 8 the most probable manifests in the previous quarter that can be used to predict whether the consumer will visit an optimetrist are depicted as a network where edge

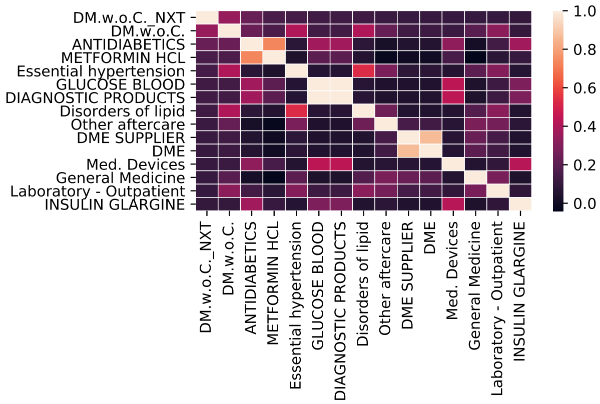

Figure 7: Feature Corrolation related to Diabetes mellitus without complication

thickness, shows the prevalence of manifests related to a specific target. As shown, those who had more medical interactions (surgery, medicines, office visit, etc.) are more probable to have optometry event in their medical journey.

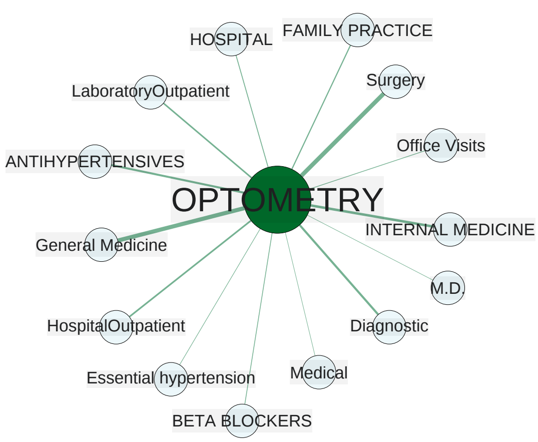

Figure 8: Optimetry targets

6 EXISTING TECHNOLOGIES

Health-care domain has specific characteristics and requirements that can not be addressed by general purpose or commercial recommender systems that are available in other domains (such as ones used in Netflix or Amazon) [5]. Recommender systems for clinical activities have no specific task and it mainly depends on the item that is recommended. All of the possible items are expected to be recommended. Rating system does not exist and most clinical behaviours are binary (have a symtom or not). Compare to the general recommender systems with specific and well defined tasks, subset of items can be recommended, rating system exists due to subjective desire and behavior is not binary as a customer may refuse to buy but it does not indicate she does not like the product. However, similar to other domains, machine learning methods have a clear advantage over manual inspection of data. Health domain is volatile and dynamic, machine learning methods can tolerate the changes in medical codes due to easier retraining process compare to manually pattern recognition, robust against temporal changes of patterns - using a sliding window approach for learning, and can offer personalized experience per hospital or patient at the same time having a general view of the whole system.

To design a framework/system, it is necessary to know the target users. Two main end users can be considered for healthcare recommender systems. Wiensner and Pfeifer [26] suggest that such systems can target health professionals (doctors and/or nurses) to help them gather additional information on a special case, or can identify patients as end user and deliver health-related content to them, such as lifestyle change recommendation [8] through changing their sleeping, eating, and exercising routines and improving patient safety [17] and lowering health risks through informing them about interactions between different drugs. Policy makers are also another target for in this domain.

To design such a system, there are several guidelines. Valdez et al. propose a 3-step process [23] to design a recommender system: 1) understanding the domain, 2) Evaluation , and 3) Inception. In the evaluation step, the importance of user-centered criteria and ethical implications (trust, value, security, long-term efficiency, individual freedom, and risk) in addition to accuracy metrics are discussed. Schafer et al. discuss the recent challenges that were tackled by the researchers and how to proceed toward a health aware recommender system. Personalization, the balance between persuasion and empowerment, and user trust and satisfaction are the main issues that captured researchers attention [19]. They group the challenges for future studies into 3 groups of Patient, recommender systems, and evaluation challenges. User modeling and profiling, Data integration and cleaning from multiple sources are the main patient-related challenges. On the recommender system side, personalized and accurate recommendation along with step by step implementation of recommendations using the "expert-in-the-loop" interactions. On the evaluation side, the accuracy, real-life performance, ethical, and privacy considerations are discussed.

There has been a number of prior efforts in this domain. One of the well known existing application is Promedas [13], a medical patient-specific clinical diagnostic decision support system, that uses a probabilistic graphical model built with the help of medical specialists. As discussed earlier, they help in recommending a diagnosis specific to an individual when there is a ambiguity among physicians without rationalization. Probabilistic methods and especially Bayesian networks have been used in a wide range of domains. For example, Huang [9] used the Bayesian network in response to two issues in the tourism domain. First, the absence of travel history for a single user to use a content-based activity estimation and the second is the absence of similarity between users and other users.

Since our system predicts multiple targets simultaneously, for each person, certain combinations of the outputs will be more likely that the other combinations. To address this fact, we adopted a structured prediction based approach that uses collective classification in its core by considering the associativity of targets as nodes in a graph [20, 24]. This model captures dependencies that would not be considered otherwise [7, 21]. At the same time, the consumers' feedback plays a vital role in fine-tuning and improving the performance of the recommender systems. However incorporating the users' feedback is challenging, the system can also be exposed to bias due to personal taste or motivations as was shown in other domains such as Amazon or application markets [12, 27]. Therefore, we incorporated robust approaches similar to Torkamani et al. and Fakhraei et al. [6, 12, 21, 22]. We also need to tackle the imposes computational complexity to the system, if the goal is to update the system's state on the fly in real time. For which, we can consider using updating schemes that have been used in similar domains to reduce the updating costs [11].

Given all the efforts have been done in this domain, our model uses a large and rich dataset and relies on a mathematically correct and proved basis and is able to address a wide range of different questions about the health journey of consumers, beneficial for consumers, providers, and payers over time.

7 CONCLUSION

Our proposed recommender system provides personalized, timely and actionable health-care insights for consumers. We make relevant suggestions by predicting the probabilities of various health events. By deploying users' feedback from their interactions within the mobile application, we enable additional personalized suggestions. This is accomplished through an ensemble algorithm, where a Bayesian network is combined with a random forest. In the future, we can improve this framework in several ways. First, by including data from other sources, such as lab results or nurse notes, we can expand the feature set to provide a more complete view of the consumer. This view would improve the precision-recall metrics of the predictions, as well as shed a brighter light on how to increase the effectiveness of the recommendations. Second, while a calendar quarter is currently the feature extraction and prediction time unit, we can change this within the probability model to predict the timing of a health event (as an additional random variable). We could also expand the model to capture a longer period of health care history to identify missing values over time, which would lead to the discovery of long-term influences, such as chronic ailments. Finally, for both better interpretability and overall improvement of the recommender system, we are working on a context-aware attention modeling algorithm to identify, invigorate, and use the most relevant features extracted from the health data and received feedback.

REFERENCES

[1] Pragati Agrawal and Amit kumar Dewangan. 2015. A brief survey on the techniques used for the diagnosis of diabetes-mellitus. *Int. Res. J. of Eng. and Tech. IRJET* 2 (2015), 1039–1043.

[2] PK Anooj. 2012. Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules. *Journal of King Saud University-Computer and Information Sciences* 24, 1 (2012), 27–40.

[3] Pushkaraj R Bhandari, Sapna P Yadav, Shyam A Mote, Devika P Rankhambe, UG Scholar, and Pune APCOER. 2016. Predictive system for medical diagnosis with expertise analysis. *International Journal of Engineering Science* 4652 (2016).

[4] Amy Compton-Phillips. [n. d.]. Care Redesign - What Data Can Really Do for Health Care. http://join.catalyst.nejm.org/hubfs/Insights%20Council%20Monthly%20-%20Files/Insights%20Council%20March%202017%20Report%20What%20Data%20Can%20Really%20Do%20for%20Health%20Care.pdf

[5] Lian Duan, W Nick Street, and E Xu. 2011. Healthcare information systems: data mining methods in the creation of a clinical recommender system. *Enterprise Information Systems* 5, 2 (2011), 169–181.

[6] Shobeir Fakhraei, James Foulds, Madhusudana Shashanka, and Lise Getoor. 2015. Collective spammer detection in evolving multi-relational social networks. In *Proceedings of the 21th acm sigkdd international conference on knowledge discovery and data mining*. ACM, 1769–1778.

[7] Shobeir Fakhraei, Bert Huang, Louiqa Raschid, and Lise Getoor. 2014. Network-based drug-target interaction prediction with probabilistic soft logic. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* 11, 5 (2014), 775–787.

[8] Robert G Farrell, Catalina M Danis, Sreeram Ramakrishnan, and Wendy A Kellogg. 2012. Intrapersonal retrospective recommendation: lifestyle change recommendations using stable patterns of personal behavior. In *Proceedings of the First International Workshop on Recommendation Technologies for Lifestyle Change (LIFESTYLE 2012), Dublin, Ireland*. Citeseer, 24.

[9] Yuxia Huang and Ling Bian. 2009. A Bayesian network and analytic hierarchy process based personalized recommendations for tourist attractions over the Internet. *Expert Systems with Applications* 36, 1 (2009), 933–943.

[10] Aiswarya Iyer, S Jeyalatha, and Ronak Sumbaly. 2015. Diagnosis of diabetes using classification mining techniques. *arXiv preprint arXiv:1502.03774* (2015).

[11] Soheil Jamshidi and Mahmoud Reza Hashemi. 2012. An efficient data enrichment scheme for fraud detection using social network analysis. In *Telecommunications (IST), 2012 Sixth International Symposium on*. IEEE, 1082–1087.

[12] Soheil Jamshidi, Reza Rejaie, and Jun Li. 2018. Trojan Horses in Amazons Castle: Understanding the Incentivized Online Reviews. In *Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '18)*. ACM.

[13] Bert Kappen, Wim Wiegerinck, Ender Akay, Jan Neijt, and Andr'e van Beek. 2003. Promedas: A clinical diagnostic decision support system. In *Proceedings of the 15th Belgian-Dutch Conference on Artificial Intelligence*. 23–24.

[14] Nguyen Cong Long, Phayung Meesad, and Herwig Unger. 2015. A highly accurate firefly based algorithm for heart disease prediction. *Expert Systems with Applications* 42, 21 (2015), 8221–8231.

[15] Joao Maroco, Dina Silva, Ana Rodrigues, Manuela Guerreiro, Isabel Santana, and Alexandre de Mendonça. 2011. Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC research notes* 4, 1 (2011), 299.

[16] Amrita Naik and Lilavati Samant. 2016. Correlation review of classification algorithm using data mining tool: WEKA, Rapidminer, Tanagra, Orange and Knime. *Procedia Computer Science* 85 (2016), 662–668.

[17] Haggai Roitman, Yossi Messika, Yevgenia Tsimerman, and Yonatan Maman. 2010. Increasing patient safety using explanation-driven personalized content recommendation. In *Proceedings of the 1st ACM International Health Informatics Symposium*. ACM, 430–434.

[18] Kanak Saxena, Richa Sharma, et al. 2015. Efficient heart disease prediction system using decision tree. In *Computing, Communication & Automation (ICCCA), 2015 International Conference on*. IEEE, 72–77.

[19] Hanna Schafer, Santiago Hors-Fraile, Raghav Pavan Karumur, Andre Calero Valdez, Alan Said, Helma Torkamaan, Tom Ulmer, and Christoph Trattner. 2017. Towards health (aware) recommender systems. In *Proceedings of the 2017 international conference on digital health*. ACM, 157–161.

[20] Ben Taskar, Vassil Chatalbashev, Daphne Koller, and Carlos Guestrin. 2005. Learning structured prediction models: A large margin approach. In *Proceedings of the 22nd international conference on Machine learning*. ACM, 896–903.

[21] MohamadAli Torkamani and Daniel Lowd. 2013. Convex adversarial collective classification. In *International Conference on Machine Learning*. 642–650.

[22] Mohamad Ali Torkamani and Daniel Lowd. 2014. On robustness and regularization of structural support vector machines. In *International Conference on Machine Learning*. 577–585.

[23] Andr'e Calero Valdez, Martina Ziefle, Katrien Verbert, Alexander Felfernig, and Andreas Holzinger. 2016. Recommender systems for health informatics: State-of-the-art and future perspectives. In *Machine Learning for Health Informatics*. Springer, 391–414.

[24] David Weiss and Benjamin Taskar. 2010. Structured prediction cascades. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. 916–923.

[25] WAJJ Wiegerinck and Tom Heskes. 2001. Probability assessment with maximum entropy in Bayesian networks. (2001).

[26] Martin Wiesner and Daniel Pfeifer. 2014. Health recommender systems: concepts, requirements, technical basics and challenges. *International journal of environmental research and public health* 11, 3 (2014), 2580–2607.

[27] Zhen Xie and Sencun Zhu. 2015. AppWatcher: Unveiling the underground market of trading mobile app reviews. In *Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks*. ACM, 10.