Tell Me What I Don't Know - Making the most of Social Health Forums

Jerome Rolia, Wen Yao, Sujoy Basu, Wei-Nchih Lee, Sharad Singhal, Akhil Kumar, Sharat Sabella
HP Laboratories
HPL-2013-43

Keyword(s):
Situational awareness; healthcare application

Abstract:
We describe a novel approach to helping patients learn about their condition using social health forums. We infer a patient's condition based on her clinical data, and assign her to a clinical state pertaining to that condition. Our system presents the patient with forum posts that are most relevant to her current state. By adjusting a LikeMe dial, the patient can modify the results from those that are most popular with patients in the same condition to the most general that apply to patients that map to other states for the condition. The posts are ranked in order of importance to the patient condition and the dial settings and the Top-N results are shown to the patient. The initial rankings of posts for different states are obtained by weights assigned by a medical expert. The patients can also give feedback with a LIKE or a DISLIKE button on whether the post is useful to them. This is incorporated into our method to dynamically change the initial weights. As the system is used in practice the initial rankings are subsumed by the feedback provided by users of the system. Our approach is general but we give results in the context of a diabetes social health forum. The algorithms are given along with preliminary results and an illustration of the user interface. We view this effort as a component of a larger patient navigation and engagement system that can help patients gain a better understanding of their condition.
Tell Me What I Don’t Know -  
Making the most of Social Health Forums

Jerry Rolia, Wen Yao, Sujoy Basu, Wei-Nchih Lee, Sharad Singhal
Hewlett Packard Labs
Palo Alto, California, USA
{jerry.rolia, wen.yao2, sujoy.basu, wei-nchih.lee, sharad.singhal}@hp.com

Akhil Kumar, Sharat R. Sabbella
Smeal College of Business
Pennsylvania State University
University Park, Pennsylvania, USA
{akhilkumar, sis5612}@psu.edu

Abstract—We describe a novel approach to helping patients learn about their condition using social health forums. We infer a patient’s condition based on her clinical data, and assign her to a clinical state pertaining to that condition. Our system presents the patient with forum posts that are most relevant to her current state. By adjusting a LikeMe dial, the patient can modify the results from those that are most popular with patients in the same condition to the most general that apply to patients that map to other states for the condition. The posts are ranked in order of importance to the patient condition and the dial settings and the Top-N results are shown to the patient. The initial rankings of posts for different states are obtained by weights assigned by a medical expert. The patients can also give feedback with a LIKE or a DISLIKE button on whether the post is useful to them. This is incorporated into our method to dynamically change the initial weights. As the system is used in practice the initial rankings are subsumed by the feedback provided by users of the system. Our approach is general but we give results in the context of a diabetes social health forum. The algorithms are given along with preliminary results and an illustration of the user interface. We view this effort as a component of a larger patient navigation and engagement system that can help patients gain a better understanding of their condition.

Keywords—health forum; patient questions; similarity; clinical state; personalized health

I. INTRODUCTION

Increasingly consumers rely on social media to make a variety of decisions about travel, shopping, service providers, etc. A rapidly growing number of consumers also use social media to stay informed about their personal health and well-being. Examples of these online sites include web portals (e.g., WebMD [1] and MedicineNet [2]), which serve as repositories of expert derived health information [3, 4], and medical forums (e.g., Diabetes Forum [5] and Diabetes Daily [6] for diabetes management), in which communities of people with shared medical issues exchange knowledge and experiences with their illnesses [7]. Unlike web portals, online medical forums offer the individual user an opportunity to ask questions or learn about specific issues that the medical expert may not be aware of. For example, recent studies suggest that many physicians are unaware of the use of alternative and complementary medicines among their patients [8], but questions about these can be found in online medical forums.

An individual’s use of online medical forums can also provide highly useful information to physicians and health care providers. The browsing history can alert the physician to those clinical issues that are important from the patient’s perspective, thereby providing an opportunity for the physician to provide expert information and context to the material that a patient reads online. However, relying a patient’s browsing history of online medical forums to the health provider poses many challenges. To begin with, the patient may need to browse through many discussion threads before finding those topics relevant to her specific condition. Many online forums offer search tools, but these search capabilities typically rely on keywords, which may provide either too many or too few results to be meaningful to the patient. Lastly, online forums are agnostic to the clinical state of the patient, and thus unable to provide information that is personalized to the patient’s health state. There is a need, then, to identify those topics in an online medical forum that are most relevant to a patient in a particular time.

We have previously described Fusion [9], a cloud based system for secure storing and sharing of electronic medical records (EMR) among health providers, patients, payers, and researchers. Fusion is a research system from HP Labs that supports the controlled exchange of private information across administrative domains, currently being built on HP Cloud Compute. In our prototype, we use Fusion as a bridge between EMR systems and a personal health record (PHR) system. The system we describe in this paper is a prototype implemented as a part of our patient portal for context-aware navigation and engagement [10], being built on top of the Fusion platform. Our contribution in this paper is that when a patient logs into Fusion, the patient can authorize the capability to deduce the state of their condition for a disease based on their PHR. Furthermore, as they browse a health forum from within Fusion, we can record their browsing behavior along with other clinical information to support the methods that are the focus of this paper. The logging of clinical information and user interaction data allows us to characterize the context based use of the forum posts without having patients reveal private information directly in the forum.

In this paper, we propose an approach to improve the online medical forum browsing experience of a patient. We accomplish this by determining the clinical state of the patient at the time she is browsing the forum, and subsequently using
that clinical state to find other similar patients who are using the forum. From there, the forum topics most relevant to the group of similar patients are presented to the user, who can then determine which topics are specifically relevant to her. The specific user topics can be shared with health providers to inform physician-patient discussions. In this way our system overcomes the limitations of other health forums that are agnostic to patient clinical state.

The organization of this paper is as follows. Section II describes our approach and the focus of the paper. In Section III we describe the sources of information and assumptions about clinical states for this paper. We use type II diabetes as a use case for illustration. After that, our method for ranking forum posts is presented in Section IV. Section V describes our prototype and results. A discussion is offered in Section VI and related work presented in Section VII. Finally, a summary and concluding remarks are offered in Section VIII.

II. APPROACH

Our proposed system must complete three tasks. First, it must determine the clinical state of the patient from the PHR at the time she is browsing an online medical forum. This enables it to determine which information from the forums will be relevant to the patient. Second, the system must determine which other users of the medical forums are similar to the patient at the time the patient browses the forum. This allows the system to classify patients based on their current clinical states and compute similarity measures between all possible pairwise clinical states. Third, the system must implement a metric that evaluates and ranks which forum topics are most relevant to the cohort of similar patients.

Figure 1 illustrates our overall approach. First, an initial set of questions \( Q = \{ Q_1, Q_2, \ldots, Q_n \} \) is obtained from a topic forum, e.g., for diabetes. Let \( S = \{ S_1, S_2, \ldots, S_m \} \) denote the set of patient states. Each selected question \( Q_i \) is tagged with a weight vector \( (w_{i1}, w_{i2}, \ldots, w_{im}) \) where \( w_{ij} \) is the weight assigned to question \( Q_i \) based on its relevance or importance to patients in state \( S_j \). The patients see the ranked questions according to the deduced state of their condition (i.e., patient state) and their preference regarding questions specific to their current state (i.e., like me weight).

To ensure that the information from the forum is returned to a patient rank ordered according to a patient’s interest, we also allow patients to explore questions by using a PatientState dial to adjust their current state, and making the questions more or less specific to their current state by adjusting a LikeMe dial. This dial specifies whether the results the patient sees should be from people who are most similar or least similar to the patient. These two dials and their controls are discussed further later in the paper.

Further, our system evolves and improves with increasing use in practice. As patients use the system, they can provide useful feedback with a LIKE or a DISLIKE button to indicate whether the suggested questions are helpful or not. They can also explicitly score the questions on a 1-5 scale. Such user feedback helps to fine-tune the rankings and eventually reach a steady state that reflects the true underlying values. Of course, patients can always post new questions to the forum, thereby adding topics and associated clinical states to the overall repository of topics.

III. PRELIMINARIES

For the rest of this paper, we describe the implementation of our system for type II diabetes mellitus. We have chosen the diabetes use case for a number of reasons. Type II diabetes mellitus is a common chronic disease in the United States, and it carries with it a significant health burden at both the individual and public health level. Additionally, several studies suggest that improved communication between the patient and health care provider and better self-management can lead to better outcomes in diabetes care [11]. An added advantage of using type II diabetes as our test case is that there exists several frequently used online medical forums dedicated to such patients, such as Diabetes Forum [5] and Diabetes Daily [6].

A. Data Sources

EMR data or broadly health records are not easily available because of privacy concerns and the effort needed to de-identify the data. The use of EMR data must comply with various regulatory policies such as HIPAA [12]. Because of the privacy issues and difficulties involved with using personalized health information from an EMR/PHR data that we generated with input from a medical professional (co-author WL). Briefly, we created a series of possible clinical pathways that an individual may go through from the time that she is initially diagnosed with diabetes. These clinical pathways are parameterized based on a collection of clinical factors, such as the measured glucose level, or the specific administration of medications. This produces an aggregate pathway that represents a sequence of clinical events based on evidence from medical literature [13].

Figure 2 illustrates a portion of these clinical pathways. The illustration includes decision nodes (labeled “D”) that govern the choice of activities and activities (labeled “A”) that correspond to interventions such as performing lab tests. Such activities often cause information to be generated that is stored in a patient’s health record. For example, activity A2 (vitals and symptoms) results in patient vital signs being generated,

![Figure 1. Overall Approach](image)
Fig. 2. Clinical Pathways for Type II Diabetes (percentages in parentheses are probabilities of a path being taken)

including weight, body temperature, blood pressure, etc. and symptoms, including fatigue, weight loss, etc.

Using the model shown in Figure 2., we simulated a population of 1000 synthetic patients, each of whom can be described by a specific path in the model. Table I shows a sample of medical records from 10 synthetic patients, their demographics, and the associated clinical path. The data shown in Table I is typical of information available from a PHR or EHR.

Next, we classified all the patients into one of these clinical states using a set of rules derived by the medical expert using medical guidelines. Rules can formally represent clinical knowledge that is used to make decisions through logical reasoning. Rules are represented as follows:

“If antecedent, then consequent”

Antecedent (rule body) can be composed of one or more conditions connected by logical relationships, such as conjunctions (AND), disjunction (OR), negation (NOT), etc. Consequent (rule head) is usually conjunctions of actions. If the conditions in the antecedent are true, then the consequent must also be true. In our examples, the consequent part of a triggered rule can be used to decide the current patient state.

Table II shows examples of how patient state assignment rules relate to the data available to us from a patient’s PHR. These rules are developed based on medical guidelines from NICE Pathways [14] and consultation with our medical expert.
For example, with rule R1, if a patient has no history of type 2 diabetes and has no incidental findings (i.e., unexpected and accidentally discovered medical problems), this patient is considered as healthy or undiagnosed (S1). Considering another example of rule R7, if a patient has history of type II diabetes, and her fasting glucose level falls out of the normal range (20, 700) or she has critical conditions, then we infer the state of this patient to be poorly controlled diabetes (S6). Here, the critical conditions can be inferred by another rule R8 from another set of medical data. New facts inferred by one rule can be used as input to potentially fire more rules. This is called forward chaining. In our system, we develop a set of rules to map patients to different states. Similar sets of rules can be developed for other diseases as well.

Table II. Sample Rules To Assign Patients To Patient Clinical States

<table>
<thead>
<tr>
<th>Rule #</th>
<th>Rule Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>IF Patient.hasT2DM==false AND Patient.hasIncidentalFindings==false THEN Patient.status = S1 (healthy or undiagnosed)</td>
</tr>
<tr>
<td>R2</td>
<td>IF Patient.hasT2DM==false AND (Patient.hasIncidentalFindings==true AND Patient.HbA1c&lt;6.5%) THEN Patient.status = S1 (healthy)</td>
</tr>
<tr>
<td>R3</td>
<td>IF Patient.hasT2DM==false AND Patient.HbA1c&gt;=6.5% THEN Patient.status = S2 (new diabetic)</td>
</tr>
<tr>
<td>R4</td>
<td>IF (Patient.hasT2DM==true) AND (Patient.fastingGlucose&lt;80 OR Patient.fastingGlucose&gt;150) THEN Patient.status = S3 (uncontrolled diabetic)</td>
</tr>
<tr>
<td>R5</td>
<td>IF Patient.hasT2DM==true AND ((Patient.fastingGlucoseE[80, 200] AND Patient.gender == female) OR (Patient.fastingGlucoseE[80, 250] AND Patient.gender == male)) THEN Patient.status = S4 (controlled diabetic)</td>
</tr>
<tr>
<td>R6</td>
<td>IF Patient.hasT2DM==true AND (Patient.hasComplications==true) THEN Patient.status = S5 (diabetic with complications)</td>
</tr>
<tr>
<td>R7</td>
<td>IF (Patient.hasT2DM==true) AND ((Patient.fastingGlucose&lt;20 AND Patient.fastingGlucose&gt;700) OR Patient.hasCriticalConditions==true) THEN Patient.status = S6 (diabetic with emergency)</td>
</tr>
<tr>
<td>R8</td>
<td>IF (Patient.hasHeartAttack==true) OR (Patient.age&gt;=80 AND Patient.hasChestPain==true) OR (Patient.age&gt;=60 AND Patient.hasHeartDisease==true) THEN Patient.hasCriticalConditions==true</td>
</tr>
</tbody>
</table>

Table III shows the mapping of the 10 synthetic patients to the six predefined states, as a result of rule-based reasoning. For example, for patient with P_ID 1, we can infer that she is in healthy or undiagnosed state as rule R1 is triggered. Similarly, rule R2 is triggered for both patients, with P_ID=2 and P_ID=4. When rule R4 is triggered, patients with P_ID=6 and P_ID=8 are inferred as uncontrolled diabetic.

C. Assigning Weights to Forum Topics

In this study, we create an example forum using posts or topics from two forums including Diabetes Forum (www.diabetesforum.com) and Diabetes Daily (www.diabetesdaily.com). Both of them are popular online communities of persons with diabetes. Information from about 100 relevant posts is selected manually. As topic titles vary widely, and because we want to demonstrate our system on a set of issues easily understandable in the setting of diabetes, we choose the topics that are phrased as questions, such as “Am I really a borderline diabetic?”. For simplicity, we do not address the content that is contained within the topic thread itself.

An initial evaluation of all our example forum topics has been conducted in this study, in terms of how they might be of interest to patients in different states. Each post has been reviewed by our medical expert, and then assigned a score for each of the six clinical states described previously. The score is an estimate of the percent of page views for patients from a given clinical state. For example, a score of 0.20 means that for that clinical state, 20% of the patients would have viewed the specific forum topic. In this way, we are able to generate an initial population model for the browsing histories of patients with diabetes.

Table IV shows a sample set of forum post questions along with their row vectors of weights (i.e. the six weight values w_i, through w_6, where w_i represents the weight of question Q_i for state S_j). These questions have been sorted by the first column. Thus, this is the order in which the questions will appear if the patient is in state S_i and is only interested in questions pertaining to that state. For example, the forum question Q_1, “Am I really a ‘borderline’ diabetic?” will be seen by only a small percentage of patients who believe they are non-diabetic, S_1 with w_1, but will be seen by a much larger proportion of persons who are newly diagnosed with diabetes, S_2 with w_2. Similarly, ketosis, which is a complication of uncontrolled diabetes, will be much more relevant to the uncontrolled diabetic states (S_5, S_6).

It should be noted that these initial, subjective question weights will be adjusted based on actual user feedback as explained in Section IV.
TABLE IV. SAMPLE QUESTIONS AND WEIGHTS FOR PATIENT STATES

<table>
<thead>
<tr>
<th></th>
<th>Question [Q]</th>
<th>W1</th>
<th>W2</th>
<th>W3</th>
<th>W4</th>
<th>W5</th>
<th>W6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Am I really &quot;borderline&quot; diabetic?</td>
<td>0.001</td>
<td>0.01</td>
<td>0.05</td>
<td>0.01</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>2</td>
<td>Do you have any problems with Ketosis?</td>
<td>0.001</td>
<td>0.1</td>
<td>0.7</td>
<td>0.3</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>3</td>
<td>How low must you keep your Blood sugar?</td>
<td>0.001</td>
<td>0.6</td>
<td>0.8</td>
<td>0.8</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td>4</td>
<td>Do people lose or gain weight with insulin?</td>
<td>0.001</td>
<td>0.2</td>
<td>0.4</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>5</td>
<td>How many times a day do you test your blood?</td>
<td>0.001</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.3</td>
<td>0.5</td>
</tr>
<tr>
<td>6</td>
<td>Why do BG levels fluctuate suddenly?</td>
<td>0.001</td>
<td>0.2</td>
<td>0.4</td>
<td>0.3</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>7</td>
<td>When is it ER time?</td>
<td>0.001</td>
<td>0.01</td>
<td>0.3</td>
<td>0.1</td>
<td>0.3</td>
<td>0.4</td>
</tr>
</tbody>
</table>

D. Clinical State Similarity Matrix

Having classified the patients into different states, we use the similarity measure between each pair of clinical states to implement the feature in which the patient dials in a scale of how relevant the forum posts should be to her specific clinical state. To derive this clinical state similarity matrix, we asked the question: How similar are two clinical states in terms of the kinds of forum topics that are of interest to each? We compute the cosine similarity of two column vectors of weights from Table IV, and use it to represent the similarity of these two states. Thus, the similarity of State $S_j$ and $S_k$ is calculated as follows, where $n$ is the total number of forum topics in our repository.

$$sim_{j,k} = \frac{\sum_{i=1}^{n} w_{ij} \times w_{ik}}{\sqrt{\sum_{i=1}^{n} |w_{ij}|^2} \times \sqrt{\sum_{i=1}^{n} |w_{ik}|^2}}$$

Table V shows the similarity matrix that includes all the possible pairwise combinations of the six clinical states we consider. This matrix has been validated by our medical expert to be reasonably close to what he would assign manually.

TABLE V. CLINICAL STATE INTEREST SIMILARITY MATRIX

<table>
<thead>
<tr>
<th>Sim(j,k)</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
<th>S6</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>1.0000</td>
<td>0.0295</td>
<td>0.0342</td>
<td>0.0311</td>
<td>0.0363</td>
<td>0.0328</td>
</tr>
<tr>
<td>S2</td>
<td>0.0295</td>
<td>1.0000</td>
<td>0.7805</td>
<td>0.8682</td>
<td>0.6689</td>
<td>0.5162</td>
</tr>
<tr>
<td>S3</td>
<td>0.0342</td>
<td>0.7805</td>
<td>1.0000</td>
<td>0.8633</td>
<td>0.8972</td>
<td>0.8346</td>
</tr>
<tr>
<td>S4</td>
<td>0.0311</td>
<td>0.8682</td>
<td>0.8633</td>
<td>1.0000</td>
<td>0.8366</td>
<td>0.7273</td>
</tr>
<tr>
<td>S5</td>
<td>0.0363</td>
<td>0.6689</td>
<td>0.8972</td>
<td>0.8366</td>
<td>1.0000</td>
<td>0.9206</td>
</tr>
<tr>
<td>S6</td>
<td>0.0328</td>
<td>0.5162</td>
<td>0.8346</td>
<td>0.7273</td>
<td>0.9206</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

IV. RANKING FORUM POSTS

With the data, clinical states, and initial topic weights in place, we turn to ranking forum posts by relevance to the clinical state of the user. To rank questions we use a metric called adjusted weight for each question $Q_j$ that calculates the weighted sum of the normalized weight of question $Q_j$ for a patient in a state $S_j$ with the sum of the normalized weights in other states $S_k$ multiplied by the similarities of those states sim$_{jk}$ as per Table V. Given that a patient is in state $S_j$, the ranking metric for question $Q_j$ is calculated as:

$$P \times \frac{w_{ij}}{\max_i (w_{ij})} + (1 - P) \sum_{k \neq j} \text{sim}_{jk} w_{ik} / \max_i (\sum_{k \neq j} \text{sim}_{jk} w_{ik})$$

where

- $\text{sim}_{jk}$ denotes the similarity between states $S_j$ and $S_k$
- $P$ denotes the weight of patient preference of questions liked by patients who are most similar or least similar to the patient (i.e., $P \in [0,1]$)
- $w_{ik}$ denotes the weight of question $Q_i$ for state $S_k$

Once the adjusted weights are calculated, we rank the set of questions by the adjusted weight such that a question with a higher value of adjusted weight is ranked higher. A weight P of 1 in the formula shows patient questions that are most highly ranked by other patients in the same state, while a P of 0 shows the patient questions that are most highly ranked by patients across all states.

The weight vector, which is initially provided by an expert, is revised as the system is used. We can ask a patient to provide additional feedback, e.g., a 1-5 score or a like/dislike selection and use it to revise a question’s weight. If a user does not provide feedback then we assume a neutral rating of 3 on a 1-5 scale. Our system uses this information to periodically re-compute weight vectors for every question.

Finally, patients can explore questions by adjusting their current state using a PatientState dial that modifies the variable $S_j$. They can adjust another LikeMe dial that modifies the variable $P$ to make the questions more or less specific to their current state. Adjustment in either parameter may result in questions being ranked differently.

V. PROTOTYPE AND RESULTS

We implemented a prototype of our system on the Fusion platform [9]. In our prototype, we exploit Fusion as a bridge between EMR systems and the patient PHR. When a patient logs into the Fusion, the patient can authorize the capability to deduce her clinical state, e.g., $S_1$-$S_6$. Furthermore, as the patient browses the prototype health forum we label the behavior with de-identified health information from the patient’s PHR. This supports the patient based ranking of posts.

In Figure 3, we see the user interface for a patient perusing a forum using our system. The patient state we deduce provides the initial setting for the PatientState dial. The patient has the option of changing the PatientState dial, either to more accurately reflect her own state, or to explore forum topics of interest to other clinical states. She also has the option to adjust the LikeMe dial to explore forum topics that are of interest to people most or least like the patient herself. The top 20 posts are illustrated in the window above the dials. In Figure 3, the PatientState dial is set to $S_2$ (newly diagnosed diabetic), and the LikeMe scale at 2 (i.e., $P=0.2$ in our ranking algorithm). As a result, the forum topics show cover issues that are of interest to many of the other states, and are not specific to the newly diagnosed diabetic. For example, the topic “How low must you keep your blood sugar?” is a topic that is relevant to states $S_1$-$S_6$. As shown in Figure 4, when the LikeMe dial is moved upward to 10 (i.e., $P=1.0$ in our ranking algorithm), the forum topics become much more relevant to a newly diagnosed diabetic. For example the topics “Is it ever possible to control diabetes by diet and exercising...”, and “Has anyone experienced side effects due to Metformin?” are concerns more specific to people newly diagnosed with diabetics. Each post also has fields that allow the patient to indicate how relevant they find the post by providing feedback with the
LIKE/DISLIKE button, or using a 1-5 scale (not shown in the screenshots). Viewing a post and providing direct feedback influences the future order of posts seen by all the patients. However, such feedback might be not immediately reflected in the current results since the overall question weights are aggregated from all users. In the long run, such changes will be visible.

Fig. 3. Ranking of Posts for Patient when PatientState=2 and LikeMe weight=0.2

Fig. 4. Ranking of Posts for Patient when PatientState=2 and LikeMe weight=1
VI. DISCUSSION

At present patients who use forums and aim to benefit from the experience of other patients like themselves must share private information online. They have to proactively look for information that related to their current situation. The system we propose is integrated with our patient portal [10] that includes PHR/EMR. Consequently, with patient permission, this makes it possible to access the private health information of the patients to determine their clinical state. De-identified information about patients can be used with labels in logs that capture the browsing behavior of patients using the forum.

Our vision is that this system will evolve and improve with increasing use. As patients use the system their feedback will help to revise and fine-tune the initial subjective expert rankings and eventually reach a steady state that reflects their true underlying values. Moreover, adding new questions will enrich the data base of information. To achieve this evolution of weights the rankings will be recomputed periodically.

Logs labeled with de-identified health information offer additional opportunities. Such viewing measures can also be used to deduce clusters (or states) of patients that have similar viewing patterns, rules that map patients to the clusters, and the viewing similarity relationships between clusters. This work is an initial step in the direction of personalized patient-centric care. As a further step with this approach, in future work we may be able to define states at a finer level of granularity to personalize them further to specific patients. We can improve this by enhancing patient similarity measures, such as those discussed in [15, 16].

VII. RELATED WORK

Recent years have seen an increasing interest in web-based Personal Health Record (PHR) systems, which allow patients to store, view, and share their medical histories, medications, lab results, etc. to better manage their health information and communicate with their healthcare providers [17, 18]. However, these systems fail to recognize and address patient concerns since engaging consumers is more than just managing and communicating their health records. A key issue to enhance diabetes self-education is to enhance patient engagement and experience, improve quality of care and promote collaborative action planning and follow-up, which can be supported by interactive media [11].

Social media have rapidly emerged as popular sources of public health information since they offer the advantages of low costs, rapid transmission through a wide community, and ease of user interaction [19]. Although the advent of social media is still relatively new, its impact on health communication and emotional support has been investigated in many studies, both in the context of general sites such as Facebook and Twitter, and health specific sites such as PatientsLikeMe and MedHelp. It is reported that “patients with diabetes, their family members and their friends use Facebook to share personal clinical information, to request disease-specific guidance and feedback, and to receive emotional support” [20]. PatientsLikeMe is unique in the sense that members can tailor questions and consultations by referring to concrete data displayed for each patient [21]. A qualitative analysis of PatientsLikeMe conducted by Frost and Massagli [21] showed that members offer support based on their own experience and advise each other on both medical issues and how to improve daily life and long term health outcomes. Other studies have devoted to learning the structure of online community [22], e.g., identifying leaders in online cancer survivor community [23]. These approaches are very useful in creating discussion, but they fail short in that they do not assist patients in terms of directly finding and ranking questions and issues that are relevant to their current state.

A 2003 survey showed that the rates of use of the internet for health care information among the general population varied across different studies from moderate to extensive [3]. Another study around the same time [8] showed that Internet use by cancer patients was quite widespread and patients felt empowered by it. Having prior access to information on the internet gave users confidence to ask questions of their doctors. It also found that “health professionals frequently fail to meet their patients’ needs fully.” A more recent 2010 study [7] of cancer patients showed that in a sample of 178 cancer listserv users, “35% chose the Internet as their preferred source of health information compared with 19% who named their oncologist.” Moreover, this study found that dissatisfaction patients were significantly more likely to rate the Internet as a better source of information than the provider.

Of course, quality varies across Internet sites that provide health information. A trust model for how patients evaluate online health information was developed by Sillence, et al. [4]. The online health information seeking behavior of people with disabilities is studied by Liang, et al. [24].

Other studies were devoted to learning the structure of online community [22], e.g., identifying leaders in online cancer survivor community [23]. Our approach empowers a community of users to place value upon the information in posts through their usage and explicit ranking behavior.

VIII. SUMMARY AND CONCLUSIONS

Medical forums are a potentially valuable source of information for patients, but the information is not organized according to the personal needs of each patient. We have described our prototype system that shows how information can be ranked and presented based on the current state of a patient’s disease. Furthermore, a patient considers posts that they may not consider otherwise either by adjusting their state or controlling how the influence of other states affects the ranking. In this way patients can, in a systematic and controlled manner, learn about relevant issues that they are not familiar with. Such issues can help them have a more effective conversation with doctors. An important future feature of this system is that it will evolve through patient interaction, and periodically revise its weight vectors and corresponding rankings dynamically based on feedback from users.

To develop and validate our prototype, we have used synthetic patient data and posts obtained from well-known health forums. A medical expert has supported the development of the techniques for generating synthetic patient data, rules for associating patients with states, and determining the initial weights for the various posts we considered. Finally,
we present the approach in the context of health forum usage by diabetes patients. Our general method allows for patient feedback to eventually influence the posts and rankings that would be viewed by patients.

Our methodology is general and it can be extended to a variety of other health concerns as well. Future work includes refining the notion of states further so that similarity across patients can be determined in a more precise manner at a lower level of granularity. We would also like to explore alternative methods to dynamically determine inter-state interest similarity. Further, it would be very useful to conduct a field study with real patients and evaluate the performance of the system.

Lastly, we believe our proposed system is a first step towards empowering patients with tools that make them more engaged in their own care. It can even inform physicians of the real-world challenges and concerns faced by the patients as they manage their disease.

ACKNOWLEDGMENT

The work of Kumar and Sabbella was supported in part by a grant from Hewlett Packard Labs.

REFERENCES