kHealth: Proactive Personalized Actionable Information for Better Healthcare

Amit Sheth
Kno.e.sis,
Wright State University
Dayton, OH, USA
amit@knoesis.org

Pramod Anantharam
Kno.e.sis,
Wright State University
Dayton, OH, USA
pramod@knoesis.org

Krishnaprasad
Thirunarayan
Kno.e.sis,
Wright State University
Dayton, OH, USA
tkprasad@knoesis.org

ABSTRACT
Mobile devices and sensors are profoundly changing the way we create, consume, and share information. Health aficionados and patients with chronic conditions are increasingly using sensors and mobile devices to track sleep, food, activity, and other physiological observations (e.g., weight, heart rate, blood pressure). This trend is leading to a paradigm shift from reactive medicine to predictive, preventative, personalized, and participatory medicine. This is also empowering an individual to more fully participate in health related decision making. To facilitate this transformation, there is a dearth of research in understanding the richness and nuances of health care data.

There are many healthcare applications that utilize mobile devices and sensors to monitor the health of an individual. With increased instrumentation such as use of smart phones and social media provides a fine-grained access to the activities of a person and population in general. Majority of analytics is focused on finding discrepancies in a single stream of observations without much insight into the problem and actionable information. kHealth analyzes observations from passive (no human involvement in data collection) and active (human input involved in data collection) sensors to provide explanations that are intelligible to individuals and when needed their clinicians for well-informed decision making.

1. INTRODUCTION
A significant portion of healthcare data used to be created in an interaction of a patient with a healthcare provider. Increasingly, we have moved away from this trend. Growth of data associated with health, fitness and well-being is now focused outside the traditional healthcare system. This growth is driven by the control and guidance of the end user (patient or anyone involved in health, fitness and well-being) complementing/corroborating traditional health data creation and collection. This trend is powered by: (a) use of Web based tools, including search and social media for health inquiry and information exchange, (b) mobile applications, and (c) growing health-centric cyber-physical systems enabled by low cost devices, many of which now communicate over Internet (termed Internet of Things (IoTs)), including those used by participants in the quantified self [3] movement. This data then characterize physical, cyber and social aspects of signals that are indicative of human health, fitness and well-being. Two important complements of these individual or personal health signals are the public health signals and population level signals. Public health signals include analysis and outcomes of data gathered by public health systems such as electronic medical records, reports from Center for Disease Control, or regional, national, or international advisory from World Health Organization. Population health signals include data that affect sections of population such as from allergy/pollen and air quality reports, as well as general conversations regarding health topics on social media platforms such as Twitter.

There is a long track record of use of telemedicine [22, 17] and a growing experience with mobile Health (mHealth) applications [18]. The key limitation of telemedicine applications is that it simply acquires data from a patient remotely. Clinicians cannot keep up with this “unintelligible” massive amounts of data, especially data collected and delivered from patients with chronic diseases as a result of continuous and long term monitoring. Current mHealth applications suffer from requiring too much active patient involvement, often asking irrelevant/redundant questions that lead to patient fatigue and disinterest. With the spread of sensors and IoTs, the growth of variety and volume of data is unabated. So what is urgently needed are analytical techniques and easy-to-use applications that can derive health-related insights from this data, to enable timely and actionable information, to ultimately improve health care.

In this paper, we present Knowledge-enabled Health (kHealth) that aims to empower an individual in their health, fitness and well-being needs. It represents an evolutionary step beyond telemedicine systems and mHealth applications. Extensive review of current state of the art in telemedicine and mHealth is beyond the scope of this paper. The key features on which kHealth sets itself apart is use of (a) relevant personal, public and population level health signals encompassing physical, cyber (Web) and social data [23], (b) semantic web technologies that include application of relevant medical knowledge in the forms of ontologies and reason-
ing, especially semantic perception and predictive analysis, and giving precedence to passive sensing over active sensing to improve patient acceptance, and (c) "intelligence at the edge by performing all computation on the users mobile device under his/her control, a strategy that is designed to address patient/individual privacy issues. Besides the research or innovation aspects, any applications that is intended for real-world use must support myriad of engineering issues, such as user experience, which we will not discuss in this paper. Instead, we will limit our attention to health applications, and not discuss potential fitness and well-being applications.

kHealth is currently being investigated for applications in (a) reducing readmission of discharged Acute Decompensated Heart Failure (ADHF) patients, (b) management of Asthma in children, for which Institutional Review Board (IRB) approvals have been obtained and early testing of kHealth kit with patients have been completed, (c) GI readmission, for which an application has been developed, and (d) geriatrics for which research and development work continues. All applications involve active clinical collaborations leading to evaluation with patients. For brevity and focus, we will limit our discussion to the Asthma application.

2. KHEALTH DIFFERENCE: CENTRAL ROLE OF BACKGROUND KNOWLEDGE

kHealth [2] is a semantic platform to enhance decision-making and improve health, fitness, and well-being, by analyzing observations from physical, cyber, and social modalities. kHealth supports contextual (e.g., condition specific) annotation, integration, and interpretation of sensor and mobile data from individuals using deep domain knowledge bases (i.e., symptom/disease knowledge bases). It also utilizes social and other data relevant to the individual’s health needs and objectives. The four primary components of kHealth include (i) Semantic Perception: the ability to convert low-level sensor observations to compact, intelligible abstractions, (ii) Knowledge-extraction from Electronic Medical Records (EMR)/Personal Health Records (PHR): the ability to enrich “incomplete” health care knowledge base (with additional causal relationships) and its use in understanding EMR/PHR, (iii) Risk Assessment: building contextual risk profiles which in turn can be used for recommending personalized actions to the individual, (iv) Intelligence at the Edge: analytical algorithms that can run on resource constrained devices, and, and (v) knowledge-based user experience (UX) improvements such as prioritizing passive sensing that does not require human involvement over active sensing that requires human involvement because excessive active sensing can dissuade human acceptance of technology such as kHealth.

kHealth utilizes background knowledge when available for interpreting health signals. Domain knowledge is used to derive abstractions and disambiguation using machine perception [16]. Machine perception generates explanations (in the form of abstractions) from observations using abductive reasoning and disambiguates among multiple candidate explanations by seeking additional inputs using deductive reasoning [13]. Figure 1 summarizes the steps in analyzing health signals from raw observational data. IntelleGO [14, 16] is a framework that integrates both abductive and deductive reasoning that can be implemented efficiently on resource constrained devices [15]. We describe the three steps for deriving actionable information below:

2.1 Data Acquisition and Aggregation

The plethora of health related information poses challenges to integration of multisensory and multimodal observations. Semantic technologies play a crucial role in this integration. First step is to filter signal from noise by separating health related observations from all other observations. Integration challenges go beyond pulling data from different sources to dealing with heterogeneous data that can require combining quantitative (precise) data from machine sensors to qualitative (imprecise) data in citizen observations. Together they provide complementary and corroborative information. Table 1 summarizes various sensors used in collecting health signals from a patient in kHealth. A practical challenge to be overcome is to make the mo-
Table 1: Health signals and their sources that will be leveraged by our algorithms for continuous, proactive, and preventive analytics for asthma management.

<table>
<thead>
<tr>
<th>Data Sources</th>
<th>Health Signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal Level</td>
<td>Physiological: Wheezometer, Nitric Oxide, Accelerometer, Microphone, Contextual Questions; Environmental: Sensor-drone, Dust Sensor, Location</td>
</tr>
<tr>
<td>Public Level</td>
<td>Everyaware, AirQuality Egg, Allergy Alerts, Social Observations (e.g., tweets), Air Quality Index, CDC, DCHCs EMR Records (periodic manual review)</td>
</tr>
</tbody>
</table>

Table 1: Health signals and their sources that will be leveraged by our algorithms for continuous, proactive, and preventive analytics for asthma management.

<table>
<thead>
<tr>
<th>Data Sources</th>
<th>Health Signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal Level</td>
<td>Physiological: Wheezometer, Nitric Oxide, Accelerometer, Microphone, Contextual Questions; Environmental: Sensor-drone, Dust Sensor, Location</td>
</tr>
<tr>
<td>Public Level</td>
<td>Everyaware, AirQuality Egg, Allergy Alerts, Social Observations (e.g., tweets), Air Quality Index, CDC, DCHCs EMR Records (periodic manual review)</td>
</tr>
</tbody>
</table>

2.2 Analysis

We outline some techniques to analyze health signals collected from personal, public, and population modalities.

2.2.1 Semantic Perception

Integrating data from multiple modalities requires managing and making sense of all the data collected over a period of time. kHealth utilizes semantic perception algorithms to transform raw data to human intelligible abstractions using medical domain knowledge. These abstractions are meant to be comprehensible to decision makers. In general, semantic perception in kHealth involves use of abductive reasoning to derive candidate explanations from sensor data, and use of deductive reasoning to disambiguate among these with the help of patient inputs and additional targeted sensor observations. This hybrid reasoning is powered by background knowledge in the form a bipartite graph that captures symptoms and diseases. Intellego in Figure 1 uses semantic perception algorithms for transforming raw data to abstractions.

2.2.2 Patient Health Score

There are a variety of sensors available for monitoring physiological parameters and this trend is going to continue with new promising devices of medical grade quality entering the market like iPhone case health monitor [11] and Scanadu [4]. While such devices are developed with supporting applications (mobile/desktop), the focus has so far been on developing sensors. Little focus has been devoted to how to convey the results of the crucial analytics of multi-modal health signals – observations spanning personal, public, and population levels in a manner that is easy to follow to be broadly effective for the patient base. We propose to develop patient health score that summarizes all the observations into a single score easily understood by a patient e.g., GREEN, YELLOW, RED zones. Patient health score can be obtained using existing domain knowledge, e.g., Asthma severity levels can be precisely expressed as some rules. In domains with no such prior knowledge, kHealth can utilize probabilistic reasoning over the data collected from patients to determine patient health score. Such scores can embody high-level actions such as consulting doctor/nurse immediately, or taking simple preventive measure (e.g., close windows) or providing reminders (e.g., did you use rescue inhaler), that can be acted upon by patients.

2.2.3 Patient Vulnerability Score

Some chronic conditions such as asthma have high symptomatic variability that depends on the environmental conditions such as air quality, pollen levels, temperature, and humidity. Patients health score is a good indicator of the current state of the patient but it can potentially depend on how vulnerable a patient is to the prevailing environmental conditions. In fact, vulnerability is highly personalized e.g., a patient may have asthma trigger as pollen while another patient may have asthma trigger as high humidity. Thus, we derive patient vulnerability score by combining current patient health score with dynamic environmental conditions. This score acts as a warning to the patients to avoid or minimize the risk of asthma attacks.

2.2.4 Personalized Actionable Information

While there may be many publicly available health related data as summarized in Table 1, we believe that personalization of health analytics is crucial for effective management of chronic conditions such as Asthma. Symptomatic variability, sensitivity to triggers, dynamic environmental conditions, varying socio economic status demands personalized attention to patients. Personalization is required at two levels: (a) risk assessment and (b) action recommendation. Figure 1 summarizes preprocessing steps for health signal understanding (risk assessment) along with action recommendation. Figure 2 summarizes interactions between various components and the nature of questions patient usually have.

Risk Assessment: Understanding factors that influence health of a person is the first step. This corresponds to Qualify in Figure 2. There may be known guidelines for categorizing various risk levels as shown in Table 2. In Table 2, number of exacerbations is 2 in 6 months for children under 5 and 2 in one year for children > 5 years of age and lung function - children < 5 years of age are not able to perform testing to assess lung function (only used) for older children. The factors that can contribute to the worsening of health are termed as risk factors. Note that risk factors may be different across patient population for the same ailment such as Asthma. After we qualify the risk factors, for a deeper insight, we need to quantify their contributions to the wors-
Figure 2: Interactions between personal and public health signals for a deeper understanding of asthma. The kHealth app combines personal, public, and population level health signals to provide a deeper understanding of the factors that contribute to the severity and control of asthma.

...enuring health condition of the patient. This corresponds to the Quantify step in Figure 2. For example, some patients may be sensitive to pollen while some others may be sensitive to smog. Further, there may be patients who may react to both pollen and smog but at different levels. Techniques for extracting probabilistic graphical model structure from data [9, 19] can be applied for identifying risk factors. Parameter learning [12, 19] for a given structure can be applied for quantifying the contribution of each risk factor.

Action Recommendation: The role of actionable information is to summarize the health of a person and convey the actions to be taken in an intuitive manner such as contact doctor/nurse (e.g., for suggestions on increasing medication dosage or frequency), or simple means for preventing/reducing exposure to triggers. Actionable information is a summary to be conveyed to patients based on risk assessment. With human in the loop, action recommendation needs personalization based on the patient context. For example, recommendation of calling a nurse may work well for most of the patients. Recording breathing rate may work only for patients who are aware of the procedure. The study of user experience of a mHealth system is a huge research topic and is out of scope of this paper.

2.4 Hybrid Reasoning for kHealth

While analyzing sensor data, a purely machine learning technique will not be able to leverage existing declarative (prior) knowledge of the domain. Declarative knowledge is expressive and reliable because it is usually human curated, and can help to deal with complexity of a domain. Specifically, this can enable inferences that can incorporate deductive or abductive reasoning, or use rule-based language. Probability is a calculus to deal with uncertainty in many domains. There have been attempts to integrate logic based formalisms with probabilistic models [8, 21, 24, 20, 12]. While there are techniques that attempt to integrate logic based formalisms with probabilistic reasoning, there is still no consensus on a natural, expressive, and efficient hybrid representation and reasoning. Our vision of hybrid reasoning is similar in motivation to Statistical Relational Learning [10] that attempts to combine logic based formalisms with probabilistic reasoning. However, we go beyond integration by using declarative and statistical knowledge as two different levels of abstraction. Specifically, we need to connect symptoms with how they get expressed in data. This is done by mapping of declarative knowledge (logic based formalisms) to its manifestation in the physical world (e.g., sensor data). This mapping has not been explored adequately by the existing works. In fact, we believe that declarative and probabilistic reasoning should be combined to solve a complex real-world problem such as Asthma management and tracking progression of Parkinsons Disease (PD). We take PD examples due to the availability of data but the same ideas apply to any other domain as well.

The representation and reasoning framework should have...
Table 2: Asthma severity levels and corresponding intensity of symptoms

<table>
<thead>
<tr>
<th>Intermittent Asthma</th>
<th>Mild Persistent Asthma</th>
<th>Moderate Persistent Asthma</th>
<th>Severe Persistent Asthma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptoms &lt; 2 days/week</td>
<td>Symptoms at least 2 days/week</td>
<td>Symptoms daily</td>
<td>Symptoms daily and all the time</td>
</tr>
<tr>
<td>No night time awakening</td>
<td>Night time awakening ≤ 2 times/month</td>
<td>Night time awakening 4 times/month</td>
<td>Night time awakening &gt; 4 times/month</td>
</tr>
<tr>
<td>Zero/One exacerbation with the use of corticosteroid</td>
<td>Exacerbation requiring oral corticosteroid 2-4 per year</td>
<td>Exacerbation requiring oral corticosteroid 2-4 per year</td>
<td>Exacerbation requiring oral corticosteroid 2-4 per year</td>
</tr>
<tr>
<td>Normal lung function</td>
<td>Lung function &gt; 80% of predicted FEV1</td>
<td>Lung function 60 - 80% predicted</td>
<td>Lung function &lt; 60% predicted</td>
</tr>
</tbody>
</table>

Table 3: Asthma control levels and recommended actions depending on the asthma severity level

<table>
<thead>
<tr>
<th>Asthma Control =⇒</th>
<th>Daily Medication Choices for starting therapy</th>
<th>Not Well Controlled</th>
<th>Poor Controlled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severity Level of Asthma</td>
<td>(Recommended Action)</td>
<td>(Recommended Action)</td>
<td>(Recommended Action)</td>
</tr>
<tr>
<td>Intermittent Asthma</td>
<td>SABA prn</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mild Persistent Asthma</td>
<td>Low dose ICS</td>
<td>Medium ICS</td>
<td>Medium ICS</td>
</tr>
<tr>
<td>Moderate Persistent Asthma</td>
<td>Medium dose ICS alone Or with LABA/montelukast</td>
<td>Medium ICS + LABA/Montelukast</td>
<td>Medium ICS + LABA/Montelukast</td>
</tr>
<tr>
<td>Severe Persistent Asthma</td>
<td>High dose ICS with LABA/montelukast</td>
<td>Needs specialist care</td>
<td>Needs specialist care</td>
</tr>
</tbody>
</table>

Figure 3: Symptom manifestation of slow movement in the Parkinson’s Disease patient CROCUS compared to the control group member APPLE

Table 4: Semantic Web Rule Language connecting symptoms to the PD severity

Person(?p)∧hasSymptom(?p,tremors)∧hasSymptom(?p,poor-balance) → hasSeverity(?p,mild-PD)

Person(?p)∧hasSymptom(?p,move-slowly)∧hasSymptom(?p,move-interruptedly)∧hasSymptom(?p,monotone-speech) → hasSeverity(?p,moderate-PD)

Person(?p)∧hasSymptom(?p,fall-prone) → hasSeverity(?p,advanced-PD)
the following desirable qualities: (1) *Open world assumption*: this is crucial in real-world problems involving partial observability (e.g., not all sensors may be available) and distributed services (e.g., pollen reporting sensor may be intermittent), (2) *Dynamism in the environment*: this is the ability to deal with uncertainty in the changing environment, (3) *User context*: this is the ability to consider personal health with public and population level health signals for personalized recommendation, and (4) *Use of domain knowledge*: this is to integrate domain knowledge in interpretation of sensor data and health signals in general. There is no single framework that can deal with all the desired qualities. We take a layered approach by representing declarative knowledge using Semantic Web representation languages and probabilistic knowledge using probability distributions. Semantic Web languages [5] (e.g., Resource Description Framework Schema - RDFS, Web Ontology Language - OWL) makes open world assumption ideal for web data which fits well to our needs. The reason for choosing Semantic Web language is that they are well studied in terms of expressiveness and complexity. There are many off the shelf reasoners with varying capabilities that work with Semantic Web representation languages. Tool support for Semantic Web language is excellent with very active developer community.

We discuss the limitation of current state-of-the-art in combining declarative knowledge with probabilistic reasoning for Physical-Cyber-Social [23] domains. There is a missing link between declarative knowledge and observations in the real-world. For example, Parkinsons Disease (PD) has symptoms such as freezing episodes, zig-zag motion, and slurred speech. But these symptoms manifest in sensor observations in a particular way. That is, we can characterize the symptoms based on the sensor data [6]. Such a mapping may be known in some well understood domains such as PD. Knowing the symptoms will not entail the knowledge of its manifestation in observations (e.g., sensor data). We analyzed the data provided by a kaggle challenge in studying PD progression [1] sponsored by The Michael J. Fox Foundation for Parkinsons Research. Figure 3 shows variation in acceleration readings for two people. APPLE and CRO-

2.5 *kHealth Mobile Application*

Asthma application for kHealth is implemented on an Android platform with a capability to collect two out of three types of health (personal and population level) signals. Sensors report their observations to an android device on which kHealth app is deployed and launched. We have explored sensors such as heart rate, blood pressure, weighing scale, sensordrone (temperature, humidity, and Carbon Monoxide), and NODE sensor (Nitric Oxide). Figure 4 summarizes the interaction of kHealth with patients, care providers, and sensors. kHealth goes beyond collecting conventional observations by posing contextually relevant questions. An example of such a question is shown in Figure 6.

Figure 5 presents few screen shots of the mobile application for Asthma. Observations are taken by patients on a daily basis and currently we are in the process of getting this data back from the patient for analysis.

3. **CONCLUSIONS**

Mobile technology and sensor proliferation has lead to great promises in healthcare for pervasive capture of health signals using IoT. These signals span physical, cyber, and social modalities demanding holistic analysis leading to various integration challenges. We presented knowledge enabled approach to health data analytics and we presented a concrete application (kHealth) that has been deployed for de-

Figure 6: kHealth system posing a contextually relevant question to the patient

CUS are dummy names of a control group member and a PD patient respectively. CROCUs acceleration readings exhibit decreased variance compared to APPLEs acceleration readings. This shows restricted motion (low amplitude acceleration readings) of CROCUS who has PD. To algorithmically characterize these variations, we may have to learn probability distributions to characterize the symptoms.

For mechanizing such an analysis, probabilistic programming [7] gives a good procedural way to explore the problem of fitting the best distribution. After characterizing the data for PD and control group, we may use the trained model to distinguish between them. While statistical approaches are good at identifying such typicalities in distributions, they have limitations when we want to perform deduction, abduction, or rule based reasoning. A declarative representation such as in Table 4 is lot more expressive and powerful than a probabilistic model (mostly at the predicate logic level) for logic based reasoning.
cision support for Asthma. We presented a novel insight of combining declarative and probabilistic models for better understanding, specifically for using high-level symptoms/features and their low-level manifestation in terms of various sensor data.

As a future work, we would like to evaluate kHealth for Asthma for its effectiveness in management and mitigation of Asthma attacks. We want to understand the deeper underpinnings of declarative and probabilistic models for building hybrid algorithms that can transcend the level of reasoning between logic based and probabilistic domains. We would like to extend our approach and evaluate it in solving other challenging problems such as readmissions in GI, Geriatrics for reducing falls, stress management and obesity.

4. ACKNOWLEDGMENTS

We thank these key clinical partners for their requirements, domain expertise and access to patients: Dr. Shalini Forbis for Asthma, Dr. William Abrahams for ADHF readmission, and Dr. Sangeeta Agrawal for GI readmission. Surendra Marupudi and Dan Vanuch have been indispensable implementation of kHealth application/kit. We also appreciate the support by Dr. Robert Fyffe, WSU’s VP of Research, for the challenge grant that partially supported this effort.

5. REFERENCES


