Personal Health Knowledge Graphs for Patients

Nidhi Rastogi, Ph.D.[1], Mohammed J. Zaki, Ph.D.[1]
[1]**Rensselaer Polytechnic Institute, Troy, New York, U.S.A.**

Abstract

Existing patient data analytics platforms fail to incorporate information that has context, is personal, and topical to patients. For a recommendation system to give a suitable response to a query or to derive meaningful insights from patient data, it should consider personal information about the patient's health history, including but not limited to their preferences, locations, and life choices that are currently applicable to them. In this review paper, we critique existing literature in this space and also discuss the various research challenges that come with designing, building, and operationalizing a personal health knowledge graph (PHKG) for patients.
Keywords : *Personal Health Knowledge Graphs, Knowledge Graphs, Diabetes.*

1   Introduction

Knowledge Graphs (KG) encode structured information of entities and their relations by capturing information retrieved from several resources. They are represented by a pre-defined ontology that uses different classes and the relationships identified between these classes. KGs confer the ability to search information efficiently, and can help find and utilize patterns in the data for improving clinical outcomes. A few examples of KGs include Google's Knowledge Graph[8], sources such as DBpedia[9]. Publicly accessible KGs have been successfully operationalized to gather insights from medical data sets. However, they do not offer reasoning over observations of daily living (ODLs), which can inform preventative care of chronic health conditions such as diabetes, alzheimer, and asthma. KGs also do not cater to results specific to a given patient. Nonetheless, progress can be made by incorporating patient lifestyle information that is contextual, personal and topical. In this regard, personal health KGs can offer solutions that encourage bringing together medical, social, behavioral, and lifestyle information and grasp nuances of a patient's health to a greater extent. Entities from a personal KG represent the daily tasks and interactions of a specific patient. For example, if a patient were to query a recommendation system for a food recipe, a KG would get called with responses that contain any food recipe recipes. On the other hand, for a query such as "What is in the lunch menu at my favorite Indian restaurant?", a recommendation system enabled by a PHKG would return more personalized responses.

2   Defining Personal Health Knowledge Graphs (PHKG)

A PHKG represents aggregated multi-modal data that includes all the relevant health-related personal data of a patient by representing it in the form of a structured graph. The data usually comes from various heterogeneous sources like smart phones, survey forms, clinical notes, and other sources that holistically capture a patient's data. There are also different terms used to describe personal KGs. Personalized KGs are curated KGs and are limited by the entities described in the general KGs. Whereas, Personal KGs complement general KGs with additional, personal information about the patient. The entities and their attributes can change with time and so will the associated data. Some of the concepts that can be used to describe a Personal Health Knowledge Graph (PHKG) are described below:

1. **Contextual** - Most of the information captured by a generic KG is not relevant to a patient. A personal health KG, instead, can represent more fluid, contextual, and more rapidly changing information about the patient. Consider a diabetes patient querying an online health platform that recommends food options to encourage healthy life style. A PHKG is cognizant to the patient's health conditions and nutritional requirements.

2. **Personal** - A PHKG infers entities and relationships that represent patient interests and information. It evolves with time based on a patient's personal preferences, and interests depending on or relating to the circumstances that form the setting for an event, statement, or idea. A PHKG considers current health condition, health goals, eating habits, food consumption, and also the cultural background of the patient.

3. **Integrated with existing Knowledge Bases** - A general purpose KG contains prominent information in the form of classes that are *globally* instantiated and accessed by most patients. Hence, the information is described

as a class or subclass and instantiated for patients. However, classes or entities that have relevance to very few patients, and are not available in the public domain can be captured by PHKGs. This kind of information is represented by small-sized, structured graphs which are integrated with the larger, general purpose KG while also ensuring that there is entity linking taking place.

Representing a PHKG - There is no standard model for representing a PHKG. So far, published research shows different models, each based on the use-cases that motivated their formation. For example, researchers[1] recommend a personal KG with a distinctive spider-web like layout where the patient is the root-node entity. We infer from the description that the patient's personal information could be represented by new classes or sub-classes to represent wide variety of entities, attributes of entities, as well as relationships between them. Another recommended[2] approach involves decomposing a large graph into sub-graphs, such that the the nodes within a sub-graph are highly inter-connected. The identification of these sub-graphs is significant as they can help uncover unknown modules in such graphs.

3 Literature Review

In this section, we review and provide a critique of various KGs approaches created to extract personal context (or similar) from patient data, especially those suffering from various chronic diseases. While use cases may vary, the intent is to offer a personalized approach for health related recommendations by utilizing small-sized PHKG. Safavi et al.[2] describe PHKGs as summaries or "mini-KG's" that contain relevant facts about the patient. Patients have limited information capacity, and not all entities are applicable to all patients. However, a limitation of this approach is that the personalized summary is static and contains limited number of entities that are pre-assigned and updated by the system. Updating these entities and relationships is unmanageable with time unless they follow a common graph pattern. Also, the initial step to create personalized summaries is the patient showing an intent, which comes when they query the recommendation system for the first time. Given the limited memory of smart phones and the processing capabilities, it is can be challenge to create PHKG 'on-the-fly' with data extracted from heterogeneous sources.

Balog et al.[1] segregate the purpose of using a KG and a PKG. In a KG (also called public KG), entities have global importance and include resources that are both general purpose and domain specific. They usually miss out on the long tail of entities not prominent enough to have their own Wikipedia articles[7] or are non-publicly available facts. However, they define personal knowledge graphs (PKG) as a the portion of the KG that the patient wants to share with the system. While this has privacy protection undercurrents, not all patients are aware enough to know what should and shouldn't be shared. Therefore, it will be ideal to build a hybrid approach where both patient and healthcare professional can collaborate and identify the requirements for the PHKG.

Gyrard et al.[3] consider a patient's personalized knowledge by gathering information from heterogeneous sources such as environmental sensors and web-based data, aggregating, managing, integrating with existing knowledge bases and meaningfully analyzing the collected information. Their kHealth[6] project assists clinicians by aggregating data that belongs to patients of asthma, obesity, and Parkinson's disease. Since the objective of this work is to make recommendations to the clinicians through knowledge gathering and data analysis, it does not support patient-side recommendations input feedback into the daily logs.

Faber et al.[4] suggest that device memory and computing resources are a limitation and should be considered as essential factors that determines the size and span of the PHKG. Smartphones today have powerful multi-core processors and large memory capacity, so their claim needs further substantiation, and and analysis to understand the bounds of memory and computation requirements.

In order to construct personal knowledge bases, Yen et al.[5] use text-based life logs shared on social media platforms such as Facebook and Twitter and are captured in the form of tweets or short statements. The purpose of this research is to provide complementary information for recall and retrieval which leads to constructing a PKG. The approach, however, falls short of entity linking with the general knowledge bases. Also, sharing one's daily life events on social media is not necessarily a reliable and consistent source of data. Health information especially may not even be shared by patients with different health conditions. Other daily event journalling too might be sparse, if at all by patients.

4 Generating a PHKG

For the most part, a brute force approach has been recommended in published research. Ontologies are created or already exist for global KGs, but not for the much smaller-sized PHKG. In order to generate a PHKG, a quick approach includes inferring patient preferences over a given general purpose KG, and then constructing the patient's PHKG from these preferences. Usually KGs already exist as part of an actual infrastructure that has collected patient data over a period of time. Faber et al.[4] recommend that once the patient is in, different approaches can be used to dynamically create and instantiate the PHKG. For instance, queries from a patient can be used as input from which other entities and relations of potential interest to the patient can be inferred and call this creating a PHKG *'on-the-fly'*. An interesting, and sometimes useful approach, it does not consider several scenarios. Such as creating relationships and identifying accurate classes for the attributes, and removal of irrelevant entities. Also, they fail to address how individual PHKG will be created or scaled, and what kind of graph structure will be used to model each PHKG.

5 Challenges

PHKG is the new research frontier towards making recommendations to patients of various chronic diseases by also incorporating daily life data. Challenges include defining the scope of health and habit related concerns to ensure personal recommendations. Also, validation mechanisms should be devised to ensure that a PHKG effectively captures a patient's personal information over an existing KG. The presence of numerous modalities to (e.g. IoT sensors, smartwatches, apps) will potentially create ingestion, processing, and scalability issues. These questions will help us drive future research work in this area. We conclude that in order to make a research leap in the area of personal health recommendation, several relevant research related questions should be answered.

References

[1] Balog, Krisztian, and Tom Kenter. "Personal Knowledge Graphs: A Research Agenda." In Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval, pp. 217-220. 2019.

[2] Safavi, Tara, Caleb Belth, Lukas Faber, Davide Mottin, Emmanuel Müller, and Danai Koutra. "Personalized Knowledge Graph Summarization: From the Cloud to Your Pocket" In 2019 IEEE International Conference on Data Mining (ICDM), pp. 528-537. IEEE, 2019.

[3] Gyrard, Amelie, Manas Gaur, Saeedeh Shekarpour, Krishnaprasad Thirunarayan, and Amit Sheth. "Personalized Health Knowledge Graph." In ISWC 2018 Contextualized Knowledge Graph Workshop. 2018.

[4] Faber, Lukas, Tara Safavi, Davide Mottin, Emmanuel Müller, and Danai Koutra. "Adaptive Personalized Knowledge Graph Summarization." In MLG Workshop (with KDD), 2018.

[5] Yen, A.Z., Huang, H.H. and Chen, H.H., 2019, July. "Personal Knowledge Base Construction from Text-based Lifelogs." In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 185-194).

[6] Sheth, A., Anantharam, P. and Thirunarayan, K., 2014. "khealth: Proactive personalized actionable information for better healthcare". In Workshop Personal Data Analytics in the Internet of Things.

[7] Lin, T. and Etzioni, O., 2012, July. "No noun phrase left behind: detecting and typing unlinkable entities". In Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning (pp. 893-903). Association for Computational Linguistics.

[8] Steiner, T., Verborgh, R., Troncy, R., Gabarro, J. and Van de Walle, R., 2012, November. "Adding realtime coverage to the google knowledge graph". In 11th International Semantic Web Conference (ISWC 2012).

[9] Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., Van Kleef, P., Auer, S. and Bizer, C., 2015. "DBpedia – A large-scale, multilingual knowledge base extracted from Wikipedia". Semantic Web, 6(2), pp.167-195.