Informatics and Control Problems 40 Issue 1 (2020)
journal homepage: www.icp.az

Modeling the knowledge from healthcare systems for machine learning applications

G.G. Abdullayeva*, N.O. Alishzade

Institute of Control Systems of Azerbaijan National Academy of Sciences, Baku, Azerbaijan

ARTICLE INFO

Article history:
Received 08.04.2020
Received in revised form 22.04.2020
Accepted 05.05.2020
Available online 21.05.2020

Keywords:
Heterogeneous knowledge
Knowledge graph
Healthcare data
Machine learning

ABSTRACT

The paper aims to prove that healthcare data analysis techniques are no longer sufficiently efficient and suitable for managing heterogeneous knowledge issues and improving healthcare data analytics due to the rapid growth and evolution of machine learning techniques. The application of artificial intelligence facilities helps surpass the human level performance and undeniably improves the diagnosis process. Due to the tremendous advancement of data acquisition in novel diagnostic devices, healthcare data is quite large and even approaching big data, which makes the application of machine learning techniques in the analysis of this data efficient. We work on the data modelling methods for the knowledge from the healthcare sector to improve applicable machine learning methods on it.

1. Introduction

Today's healthcare industries are moving from volume-based business into value-based business, which requires an overwork from doctors and nurses to be more productive and efficient. This will improve healthcare practice, changing individual life style and driving people to longer life, prevent diseases, illnesses and infections. Over the last few years, healthcare data has become more complex because large amount of data is available, along with the rapid change of technologies and mobile applications and new diseases have been discovered [1]. Therefore, healthcare sectors believe that healthcare data analytics tools are a really important subject to manage a large amount of complex data, which can lead to improved healthcare industries and help medical practice to reach a high level of efficiency and work flow accuracy, if these data analytics tools applied correctly, but the questions are how healthcare organizations are applying these tools today, and how to think about its future use? Also, what are the challenges they face when using such tools? And finally, what innovations can healthcare add to meet these challenges?

The healthcare sector is already an early adopter and beneficiary of technological advances. Nowadays, machine learning (a subset of artificial intelligence) plays a key role in many health-related realms, including the development of new medical procedures, the handling of patient data and records, and the treatment of chronic diseases. Despite that healthcare data was small until the last few years, according to evariant.com the amount of healthcare data available is expected to reach roughly 25,000 petabytes by 2020. By using machine learning to derive insights from patterns and correlations found in healthcare data, healthcare marketers can make predictions about which patients

*Corresponding author.
E-mail addresses: ag_gulchin@rambler.ru (G.G. Abdullayeva), nigar.alish@isi.az (N.O. Alishzade)

www.icp.az/2020/1-08.pdf

may have a propensity toward certain conditions. Techniques of ML and AI have played an essential role in the medical field like medical image processing, computer-aided diagnosis, image segmentation, tumor detection, attack prediction, clinical decision support, prophylactics recommending, and etc.

The aforementioned use cases share one common aspect: they are all about images or signals that can be easily handled by any data analysis system. When it comes to unstructured data, more advanced data model is supposed to be designed and implemented. Relying on the previous researches on the subject [2], we assert that the knowledge graph is a suitable model for this purpose, due to its capability of expressing heterogeneous knowledge in various domains, in as usable form as possible, and fitting as many use cases as possible.

2. Information system and information technology in the healthcare sector

The healthcare sector is widely considered as one of the most important industries in information technology (Wager 2005). Information technology has been increasingly considered as a practice that facilitates healthcare performance through using data and information efficiently within the healthcare sector [3]. Therefore, Wager et al (2005) said that in order to understand the relation between information technologies and healthcare, we first need to understand what technologies are used in healthcare. Information technology functions have developed over the last few years not only as a technology services provider, but also as a strategic provider that develops and integrates industries' infrastructures to facilitate and ensure quality of service (LeRouge et al 2007). In the mid-80's, information technology changed the healthcare industry and brought many benefits when they used microcomputers, which were small in shape, fast and very powerful for that time. Moreover, this allowed hospitals to develop clinical applications for various medical care settings. As a result, hospitals started to purchase and adopt information systems in the healthcare industries, and after that, challenges began to emerge when professionals tried to integrate data among these systems (Wager et al 2005). However, Bhattacherjee and Hikmet (2007) and Castro (2007), granted that information technology has improved healthcare industries, but they also highlighted some of the difficulties related to the use of information technology in healthcare sectors, as they noticed that it is hard to implement information technology in small clinics and organizations, with high costs due to reduced efficiencies of scale. Therefore, IT implementation requires long term training and retention of skilled professionals. On the other side of the debate, (Abbott and Coenen 2008) they believe that information systems and information technology occupy a high position in improving healthcare industries in general, and in electronic healthcare record (HER) in particular; for the reason that implementing such technologies can save costs and times associated with daily hospital data records, such as patients schedules and billing. This is in addition to improving healthcare performance and efficiency by eliminating manual data records and paper work, and alongside smooth and flexible tracking of patient details [4, 5].

3. Healthcare analytics and data mining

Data Mining is described as a process by which data is gathered, analyzed and stored in order to produce useful and high quality information and knowledge [6]. This term also includes the way of how this data is gathered, filtering and preparation of the data for use and finally the processing of data to support data analytics and predictive modelling (Russom 2011).

The first stage of data mining is the process of gathering and collecting data [7]. However, even before gathering the data, ideas and plans should be assumed to decide which data should be gathered in order to collect specific data as desired and use it efficiently (Lamont, 2010). Furthermore, Chordas (2001) added that a lot of projects fail and exceed estimated costs because of poor quality of gathered

data which can result from poor data cleaning.

Data mining has also been applied [8] to analyze the information seeking behavior of health care professionals, and to assess the feasibility of measuring drug safety alert response from the usage logs of online medical information resources. Researchers analyzed two years of user log-in data in UpToDate website to measure the volume of searches associated with medical conditions and the seasonal distribution of those searches. In addition, they used a large collection of online media articles and web log posts as they characterized food and drug alert through the changes in UpToDate search activity compared to the general media activity. Some researchers [9] examined changes of key performance indicators (KPIs) and clinical workload indicators in Greek National Health System (NHS) hospitals with the help of data mining. They found significant changes in KPIs when necessary adjustments (e.g., workload) were made according to the diagnostic related group. The results remained for general hospitals like cancer hospitals, cardiac surgery as well as small health centers and regional hospitals. Their findings suggested that the assessment methodology of Greek NHS hospitals should be re-evaluated in order to identify the weaknesses in the system, and improve overall performance. And in home healthcare, another group of researchers [10] reviewed why traditional statistical analysis fails to evaluate the performance of home healthcare agencies. The authors proposed to use data mining to identify the drivers of home healthcare service among patients with heart failure, hip replacement, and chronic obstructive pulmonary disease using length of stay and discharge destination.

4. Healthcare data and big data analytics

One of the most important elements in dealing with and managing data is to know where and how this data will be stored once when it is collected. The traditional methods of storing and retrieving such data are not efficient anymore, since it was structured and stored in data warehouses and relational databases, after extracting and loading it from different outside sources. However, this data is transformed and classified before being ready to use and function (Bakshi 2012). Furthermore, Herodotou et al (2011) agreed with Bakshi (2012) when he said that there are many numbers of data sources now and that a huge amount of data has become available, so this growth of data will absolutely require an agile database which can deal with the data logically and through data synchronization in order to adapt to the rapid data evolution. On the other hand, Plattner and Zeier (2011) stated that databases only manage server memory data, therefore eliminating the option of managing other storage devices such as: disk and compact drivers. Accordingly, this will reduce the efficiency of database performance and real time response during the time.

5. Machine learning and healthcare data

Using data, machine learning has driven advances in many domains including computer vision, natural language processing (NLP), and automatic speech recognition (ASR) to deliver powerful systems (e.g., driverless cars, voice activated personal assistants, automated translation). Machine learning's ability to extract information from data, paired with the centrality of data in healthcare, makes research in machine learning for healthcare crucial. Interest in machine learning for healthcare has grown immensely, including work in diagnosing diabetic retinopathy, detecting lymph node metastases from breast pathology, autism subtyping by clustering comorbidities, and largescale phenotyping from observational data. Despite these advances, the direct application of machine learning to healthcare remains fraught with pitfalls. Many of these challenges stem from the nominal goal in healthcare to make personalized predictions using data generated and managed via the medical system, where data collection's primary purpose is to support care, rather than facilitate subsequent

analysis. Existing reviews of machine learning in the medical space have focused narrowly on biomedical applications, deep learning tasks well suited for healthcare, the need for transparency, and use of big data in precision medicine. Here, we emphasize the broad opportunities present in machine learning for healthcare and the careful considerations that must be made. We focus on the electronic health record (EHR), which documents the process of healthcare delivery and operational needs such as tracking care and revenue cycle management (i.e., billing and payments). While we choose to focus on the inpatient setting as the majority of machine learning projects currently focus on this data-rich environment, we note that clinical data is heterogeneous, and comes in a variety of forms that can be relevant to understanding patient health.

In tackling healthcare tasks, there are factors that should be considered carefully in the design and evaluation of machine learning projects: causality, missingness, and outcome definition. These considerations are important across both modeling frameworks (e.g., supervised vs. unsupervised), and learning targets (e.g., classification vs. regression). Many of the most important and exciting problems in healthcare require algorithms that can answer causal, “what if?" questions about what will happen if a doctor administers a treatment. These questions are beyond the reach of classical machine learning algorithms because they require a formal model of interventions. To address this class of problems, we need to reason about and learn from data through the lens of causal models. Learning from data to answer causal questions is most challenging when the data are collected observationally; that is, it may have been influenced by the actions of an agent whose policy for choosing actions is not known. In healthcare, learning is done almost exclusively using observational data, which poses a number of challenges to building models that can answer causal questions. For instance, Simpson's paradox describes the observation that the relationship between two variables can change directions if more information is included in the model. To better understand this issue, consider prior work in which researchers found that asthmatic patients who were admitted to the hospital for pneumonia were more aggressively treated for the infection, lowering the subpopulation mortality rate. A model that predicts death from asthma will learn that asthma is protective. If, however, an additional variable to account for the level of care is included, the model may instead find that having asthma increases the risk of death. This example demonstrates that causal models are not only useful to evaluate treatments, but can also help to build reliable predictive models that do not make harmful predictions using relationships caused by treatment policies in the training data. To account for these challenges, strong assumptions must be made that cannot be statistically checked or validated; i.e., gathering more data will not help.

Even if all important variables are included in a healthcare dataset, it is likely that many observations will be missing. Truly complete data is often impractical due to cost and volume. Learning from incomplete, or missing, data has received little attention in the machine learning community, but is an actively studied topic in statistics. Because healthcare is a dynamic process where vitals are measured and labs are ordered over time by doctors in response to previous observations there are strong dependencies between what variables are measured and their values, which must be carefully accounted for to avoid biased results. There are three widely accepted classifications of missing data mechanisms; i.e., the measurement mechanism determining whether a value is recorded or not. The first, missing completely at random (MCAR), posits a fixed probability of missingness. In this case, dropping incomplete observations, known as complete case analysis, is commonly used (albeit naively), and will lead to unbiased results. Second, the data may be missing at random (MAR), where the probability of missingness is random conditional on the observed variables. In this case, common methods include re-weighting data with methods like inverse probability of censoring weighting or using multiple imputations to in-fill. Finally, data may be missing not at random (MNAR), where the probability of missingness depends on the missing variable itself, or other missing and unobserved variables. Sources of missingness must be carefully understood. Sources of missingness should be carefully examined before deploying a learning

algorithm. For example, lab measurements are typically ordered as part of a diagnostic work-up, meaning that the presence of a datapoint conveys information about the patient's state. Consider a hospital where clinical staff measures patient lactate level. If a power outage led to a set of lactate levels being lost, the data are MCAR. If nurses are less likely to measure lactate levels in patients with traumatic injury, and we record whether patients were admitted with trauma, the data are MAR. However, if nurses are less likely to measure lactate levels when believed to be already, then the lactate measures themselves are MNAR, and the measurement of the signal itself is meaningful. The key feature of missing data is that there may be information conveyed by the absence of an observation, and ignoring this dependence may lead to models that make incorrect, and even harmful, predictions. Include missingness in the model. Including missingness indicators provides the most information for making predictions. However, learning models without an appropriate model of missingness leads to issues such as inaccurate assessment of feature importance and models that are brittle to changes in measurement practices. For example, troponin-T is commonly measured only when a myocardial infarction is considered probable. A model learned by treating troponin-T as MCAR would likely overpredict the rate of myocardial infarction on data where troponin-T was more regularly measured. A model trained with MAR troponin values would be more robust to this. Missingness can reflect human biases. We note that data may also be missing because of differences in access, practice, or recording that reflects societal biases. Models trained on such data may in turn exhibit unfair performance for some populations if a machine learning practitioner is unaware of this underlying variation; thus, checking models across groups is important.

In most existing work, models are trained on the largest dataset possible and assumed to be fit for deployment, i.e., models do not keep learning. This is problematic in clinical settings, because patient populations and recommended treatment procedures will change over time, resulting in degraded predictive performance as the statistical properties of the target change. For example, clinicians previously assumed that estrogen was cardioprotective in menopausal women and hormone therapy was routinely prescribed as a preventative measure until large trials reported either no benefit or an increase in adverse cardiac events. In developing new models for healthcare, models must be made robust to these changes, or acknowledge their mis-calibration for the new population. Internal Validity - Shift over time. In a notable example of concept drift, Google Flu Trends persistently overestimated flu due to shifts in search behaviors. The initial model was a great success, leveraging Google search data to predict flu incidence; however, without update the model began to overestimate flu incidence in subsequent years as user search behaviors had shifted. While the drift was unintentional, the example motivates the need for models that continuously update. External Validity - Shift over sources. There is also no reason to believe a priori that a model learned from one hospital will generalize to a new one. Many factors impact generalizability, including local hospital practices, different patient populations, available equipment, and even the specific kind of EHR each uses—transitions from one EHR to another create non-obvious feature mapping problems [11]. This issue will remain until infrastructure to easily test across multiple sites becomes prevalent. The absence of such standardization creates opportunities with respect to data normalization and the development of models that are robust to differences in data collection at different sites [12]. Creating models robust to feedback loops. Models that learn from existing clinical practice are susceptible to amplifying the biases endemic to modern healthcare. While not yet observed in healthcare, such feedback loops have been noted in the deployment of predictive policing. Such biases reflected in deployed predictions can propagate into future training data, effectively creating a feedback loop that causes further bias. Work in algorithmic fairness should be considered as it motivates the need for systems that are sufficiently aware that they can alert us to such unwanted behavior. Additionally, models trained on EHR will learn from existing procedures, rather than represent formally best-practice protocols, which may be problematic given low manual compliance in much clinical practice.

6. Knowledge modelling

Since in practice the biomedical knowledge usually comes from different domains and has different types we can assume it as heterogeneous knowledge. Concretely, the term heterogeneous knowledge uses to refer to: entities (things), their relations, and their attributes. For instance, in a healthcare database, we may find entities such as patients, medical workers, hospitals, health insurance companies, etc. Relations express which patient takes control of a particular doctor, which hospital a doctor works for, and so on. Attributes can be simple strings such as names and insurance numbers, but also richer media like short biographies, photographs like a medical card.

Although, no data model satisfies all use cases, and the knowledge graph is no exception. For example, in a simple image classification task, it would not be efficient to encode pixels of magnetic resonance or ultrasound images as separate entities in a knowledge base. But we can determine the images themselves as entities, with the raw data of a medical expert's commentaries as their attribute. Furthermore, these attributes can play a role of labels in the supervised machine learning process on the next stages of building of a healthcare system. This approach provides a data model and machine learning pipeline that allows to extend datasets with other knowledge will have been gained during the next stages of a patient's condition.

7. Working on knowledge graphs

Increasingly large electronic health records (EHRs) provide an opportunity to algorithmically learn medical knowledge. A causal health knowledge graph could learn relationships between diseases and symptoms and then serve as a diagnostic tool to be refined with additional clinical input. Prior research has demonstrated the ability to construct such a graph from over 270,000 emergency department patient visits. Moreover, clinicians are often interested in the causal relationship between diseases and symptoms. Given presenting symptoms, what is the likely diagnosis? An accurate understanding of these relations can be leveraged to build diagnostic tools, which have been shown to be useful for training clinicians, assisting clinical workflows, or even in substitution of clinicians [13].

The increased availability of electronic health knowledge records allows researchers to readily learn latent patterns from observational data. In contrast to clinical trials which are conducted on restricted subpopulations, the breadth of electronic health records (EHRs) allows for the inclusion of all patients who enter the healthcare system. With this data, researchers can build models to extract general medical knowledge and diagnose patients [14].

Although diagnostic knowledge exists in medical textbooks or online repositories like Mayo Clinic, inferring that medical knowledge automatically from EHRs provides different strengths. First, because EHRs provide a dramatically different perspective than idealized and curated medical textbooks, we may find new connections between diseases and symptoms. Second, the automated nature of extraction allows for any large EHR system to be used as source dataset. Armed with medical knowledge gleaned from EHRs and canonical medical training, clinicians can leverage both for improved clinical decision making. However, the use of EHRs to build health knowledge graphs is not without potential bias. Because of the nature of data collection, models learned on EHRs like health knowledge graphs are subject to many sources of statistical bias. First, narrow sample sizes of subsets of the data can cause underfitting despite the larger scale of the entire dataset. Second, confounders not measured by the data may bias the reliability of resulting models. Lastly, algorithms or findings from algorithms may not generalize to entirely different populations. It is essential that we closely examine so-called health knowledge graphs so that they can be used as the first steps of a diagnostic tool to improve clinical workflow and better understand diseases. We seek to understand for which diseases and for which patients a health knowledge graph performs poorly and understand

potential confounders. Error analysis can guide steps to improve derived models. For example, a change in model formulation could improve performance if sources of error are well understood. Moreover, data augmentation through additional features or a more broadly collected dataset could provide additional signal to the health knowledge graph. We hope that this critical evaluation of extracted health knowledge graphs provides an example on how to assess the robustness of medical knowledge extracted from EHRs.

8. The main challenges

Our goal is to explore an efficient way to organize, integrate, and deliver the heterogeneous tremendous medical knowledge using semantic web technologies. Therefore, there are mainly three challenges:

(1) A model is needed to organize and integrate the heterogeneous medical information. Health data from Electronic Health Records (EHRs) systems are always highly complex. It contains a mixture of many continuous variables and a large number of discrete concepts. Most of them are represented as unstructured free-text format that need nature language processing. In addition, healthcare-related terminologies may vary from different doctors [15]. As well as the health data, medical knowledge also faces similar problems, such as multiple heterogeneous variables, unstructured free-text format, and inconsistent terminology usage. Therefore, we need to propose a model to deal with this heterogeneous medical information. Moreover, in order to make computers understand this information, the conceptual graph based knowledge representation methods must be taken into consideration.

(2) To automatically retrieve knowledge from heterogeneous textual knowledge sources, effective algorithms are required to process these textual medical knowledge as the model represented.

(3) For the delivery of reasonable health knowledge, an inference algorithm is needed when we perform query and inference over the graph knowledge base.

9. Methodology

In order to organize and integrate the heterogeneous healthcare information, we propose a Healthcare Information Organization Model to normalize the heterogeneous healthcare information into a sharable and consistent format. To enhance semantic applicability, we model those information using conceptual graph representation. Overall architecture of health information system is illustrated in Figure 1.

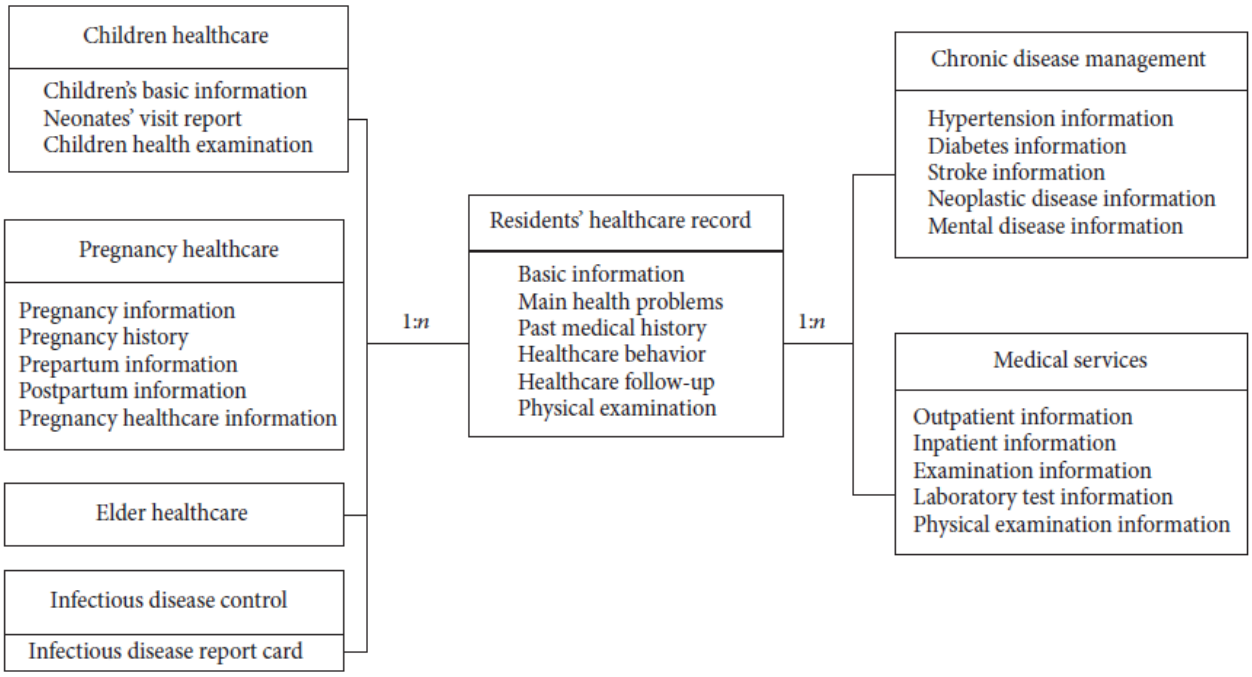

Figure 1. Overall architecture of health information system.

Medical Knowledge Model (MKM) is used to define the schema of knowledge to represent the Text Medical Knowledge (TMK) into conceptual graphs and to integrate with health data. In order to enable computers to explicate medical knowledge, we abstracted the textual format medical knowledge into a graph expression based on the conceptual graph knowledge representation [16]: medical terminologies are classified and served as the vertexes (entities) of the graph, and sentences that describe relationships between medical terms are abstracted as the verges of the graph. In addition, the descriptive knowledge which explains the entities is taken as the attributes of the entities. This meta knowledge composes the basis of our graph knowledge base. Figure 2 illustrates the graph representation of encyclopedia on pneumonia. Based on the graph knowledge representation, our MKM defines the classes (or concepts) of the medical entities with their relationships of medical knowledge that needed to be abstracted and integrated. Entities of concepts in MKM are defined in the Terminology Glossary. In order to illustrate the complicated semantics and relationships in the knowledge model, we adopt ontology technique to represent the MKM. Actually, there are many existing knowledge models in biomedical domain. Most of those knowledge models focused on a specific domain. For example, the OBO foundry [17] has developed many biomedical ontologies that are both logically well-formed and scientifically accurate. The SemanticHealthNet [18] project also developed several biomedical knowledge models for sharing knowledge. Such knowledge models can be considered and reused to build the MKM. The existing domain-specific knowledge models can be integrated through the MKM.

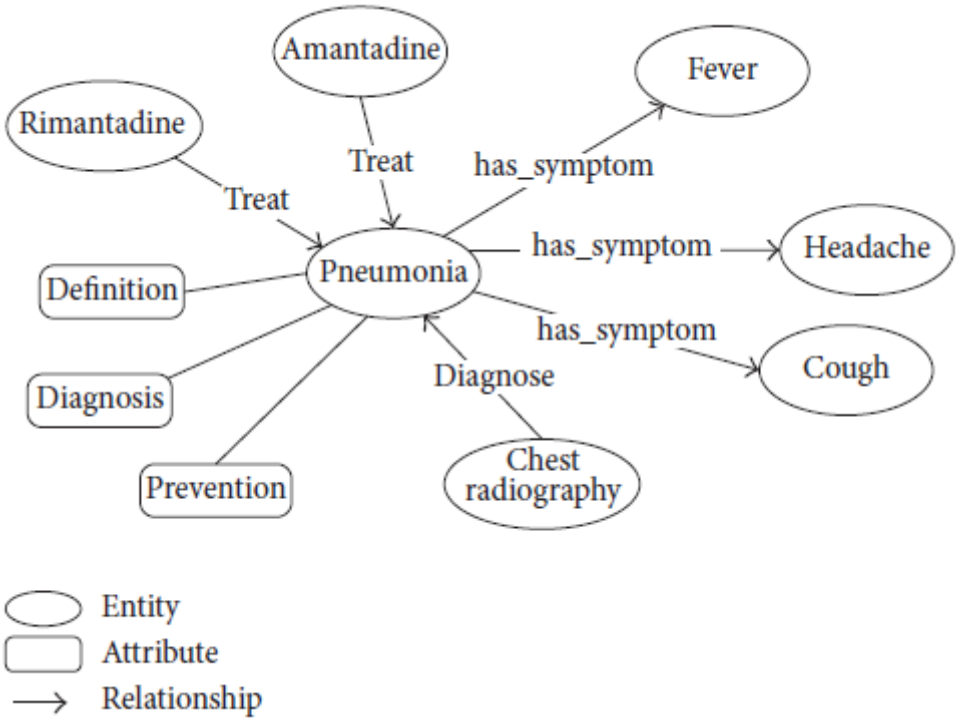

Figure 2. Illustration of the conceptual graph knowledge representation of encyclopedia on pneumonia.

In order to integrate the heterogeneous health data with medical knowledge, it is necessary to express these data into a sharable and consistent format. Fortunately, numerous studies have noticed this problem. The semantic web provides a common framework that allows data to be shared and reused across applications, enterprises and community boundaries, and receives widely adopted in healthcare data integration [19]. Hence, we adopt semantic technologies to achieve the integration of health data with medical knowledge. Health Data Model (HDM) is derived from the original data schema and supervises the health data into semantic format. We use an ontology model to express the HDM. The normalized health data tuples are stored in RDF to integrate with medical knowledge, as illustrated in Figure 3.

```
<?xml version="1.0"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:ehr="http://example/ehr#">
<rdf:Description
  rdf:about="http://example/ehr/ID000001">
    <ehr:patient>Bob</ehr:patient>
    <ehr:chief_complaint>blurred vision in left eye</ehr:chief_complaint>
    <ehr:symptom>blurred vision</ehr:symptom>
    <ehr:diagnosis>Herpes simplex keratitis</ehr:diagnosis>
.
.
.
</rdf:Description>
</rdf:RDF>
```

Figure 3. Illustration of RDF representation of EHR.

Since the health data we retrieved are stored in a relational database from EHR systems, their logical structures are defined using entity-relationship models (ERM). As a consequence, we transform the ERM to ontological model using the following steps:

(1) Identify the health data that need to integrate with knowledge.
(2) For the unstructured health data, build the structural ontological model of health data based on the existing standard.
(3) After the health data are wholly structuralized, give the detailed definition of the data domain and attributes.

Based on the above steps, our HDM is depicted as an ontology in Figure 4.

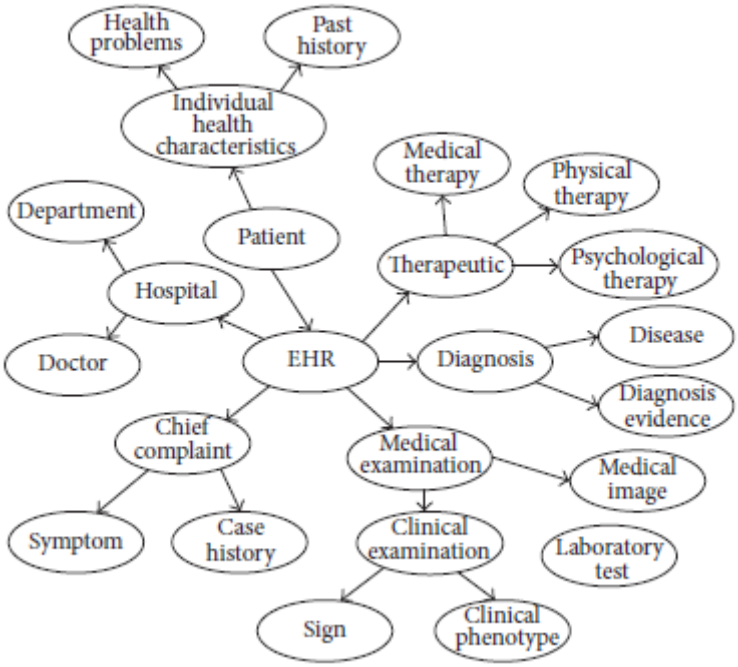

Figure 4. Illustration of health data model as an ontology.

10. Conclusion and future work

The tremendous amount of medical knowledge which emerged in recent years provides us with an opportunity to share and utilize these medical knowledge together to explore and get access to valuable useful information. We propose a contextual inference pruning algorithm to explore complex semantics between entities in chain inference while pruning meaningless inference chains. Finally, we implement our method using semantic techniques, and two prototypes are implemented to show the semantic applications on the integration of tremendous heterogeneous healthcare knowledge.

Nowadays in medicine, accurate diagnosis of a disease depends upon the successful application of AI and ML techniques. In modern machine learning, raw data is the preferred input for most statistical models. Where in the previous approaches biostatisticians were engineering features, manually picking out the important details, they now prefer data in the raw form, which are of different types and come from different domains. Moreover, crafting feature vectors which can be used as input for learning algorithms as data scientists do in traditional machine learning processes, includes adding, removing, and reshaping data, and can cause the loss of information and therefore, accuracy. If we choose a data model suited to represent this knowledge, we can develop end-to-end learning models that can directly consume knowledge graphs. We suggest that further work along these tracks should aim the development of deep learning models on knowledge bases that will enable us to provide insights out of heterogeneous healthcare knowledge from different perspectives.

In this paper, we figured out that heterogeneous knowledge coming from healthcare systems

needs a different data model than the relational model. We focused on knowledge graphs as a suitable model and created an ontology that helps understand the depth of benefits of the application of ontologies in medical information systems. Then we justified the suitability of this model for applying machine learning techniques on it and grasping more insights towards medical diagnosis and prognosis.

We believe the most promising areas of future research involve incorporating additional datasets across disparate medical settings in order to builds clinician trust as well as expose area for potential improvement. Health knowledge graphs can be used to provide automated recommendations for clinicians as well as find new edges between diseases and symptoms to advance understanding of disease.

References

[1] J.-J. Yang, J. Li, J. Mulder, Y. Wang, S. Chen, H. Wu, Q. Wang, H. Pan, Emerging information technologies for enhanced healthcare, Comput. Ind. 69 №5 (2015) 3-11.

[2] X. Wilcke, P. Bloem, de V. Boer, The Knowledge Graph as the Default Data Model for Machine Learning; IOS Press: Amsterdam, The Netherlands, 2017.

[3] J.W. Cortada, D. Gordon, B. Lenihan, The Value of Analytics in Healthcare, IBM Institute for Business Value; Armonk, NY, USA: 2012. Report No.: GBE03476-USEN-00.Conference Name:ACM Woodstock conference

[4] H.-U. Prokosch, T. Ganslandt, Perspectives for medical informatics, Methods Inf. Med. 48 №1 (2009) 38-44.

[5] A.F. Simpao, L.M. Ahumada, J.A. Gálvez, M.A. Rehman, A review of analytics and clinical informatics in health care, J. Med. Syst. 38 №4 (2014) 45.

[6] D. Tomar, S. Agarwal, A survey on Data Mining approaches for Healthcare, Int. J. Bio-Sci. Bio-Technol. 5 №5 (2013) 241-266.

[7] M. Herland, T.M. Khoshgoftaar, R. Wald, A review of data mining using big data in health informatics, J. Big Data. 1 №2 (2014).

[8] A. Callahan, I. Pernek, G. Stiglic, J. Leskovec, H.R. Strasberg, N.H. Shah, Analyzing information seeking and drug-safety alert response by health care professionals as new methods for surveillance, J. Med. Internet Res. 17 №8 (2015) e204.

[9] A. Christodoulakis, H. Karanikas, A. Billiris, E. Thireos, N. Pelekis, "Big data" in health care Assessment of the performance of Greek NHS hospitals using key performance and clinical workload indicators, Arch. Hellenic Med. 33 №4 (2016) 489-497.

[10] E.A. Madigan, O.L. Curet, A data mining approach in home healthcare: Outcomes and service use, BMC Health Serv. Res. 6 №1 (2006) 18.

[11] AK. Manrai, G. Bhatia, J. Strymish, IS. Kohane, SH. Jain, Medicine's uncomfortable relationship with math: calculating positive predictive value, JAMA internal medicine. 174 №6 (2014) 991-993.

[12] A. Subbaswamy, S. Saria, Counterfactual normalization: proactively addressing dataset shift using causal mechanisms, In: Uncertainty in Artificial Intelligence (UAI). (2018) 947-957.

[13] A.Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M.P. Turakhia and A.Y. Ng, Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network, Nature medicine. 25 (2019) p.65.

[14] S.G. Finlayson, P.LePendu and N.H. Shah, Building the graph of medicine from millions of clinical narratives, Scientific data. 1 (2014) p.140032.

[15] K.J. Cios and G.W. Moore, "Uniqueness of medical data mining," Artificial Intelligence in Medicine. 26, №1-2 (2002) 1-24.

[16] M. Chein and M. Mugnier, Graph-based Knowledge Representation: Computational Foundations of Conceptual Graphs, Springer, Berlin, Germany, 2008.

[17] B. Smith, M. Ashburner, C. Rosse et al., "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration," Nature Biotechnology. 25 №11 (2007) 1251-1255.

[18] TheSemanticHealthNet Project, 2016, http://semantichealthnet.eu.

[19] C. Tao, J. Pathak, and S. R.Welch, "Toward semantic web based knowledge representation and extraction from electronic health records," in Proceedings of the 1st International Workshop on Managing Interoperability and Complexity in Health Systems (MIXHD '11). (2011) 75-78.