A Work Project, presented as part of the requirements for the Award of a Master Degree in Management from the NOVA – School of Business and Economics.

INCORPORATING COMPLEX DOMAIN KNOWLEDGE INTO A RECOMMENDER SYSTEM IN THE HEALTHCARE SECTOR

JOEL DE JESUS FEITEIRA E PEITO – 33666

Project carried out on the Master in Management Program, under the supervision of: Qiwei Han

08-09-2020

Abstract:

In contrast to other domains, recommender systems in health sector may benefit particularly from the incorporation of medical domain knowledge, as it provides meaningful and personalised recommendations. With recent advances in the area of representation learning enabling the hierarchical embedding of health knowledge into the hyperbolic Poincaré space, this thesis proposes a recommender system for patient-doctor matchmaking based on patients' individual health profiles and consultation history. In doing so, a dataset from a private healthcare provider is enriched with Poincaré embeddings of the ICD-9 codes. The proposed model outperforms its conventional counterpart in terms of recommendation accuracy.

Keywords:

Healthcare Analytics, Machine Learning, Recommender Systems, Poincaré Embeddings, Primary Care, International Classification of Diseases

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

I. Introduction

With the emergence of healthcare analytics and growing calls to leverage the potential of the tremendous amounts of available electronic healthcare records from practitioners and scholars alike, machine learning (ML) solutions, such as recommender systems (RS) experience growing relevance in the healthcare industry (Ghassemi et al., 2018). In fact, patients do increasingly seek bespoke and digital medical solutions, similar to what they are used to from e-commerce and other domains. However, as patients' relationship to their doctors can be very personal and health is a sensitive topic in itself, healthcare recommender systems (HRS) are subject to a different set of rules and evaluation criteria, than other commercial applications of RS. For instance, product or movie RS do not operate under the same scrutiny regarding the reliability and trustworthiness of their predictions, since the ramifications of specific treatment or doctor recommendations are obviously more severe in nature.

In general, RS often capitalise on the target user's interaction data without the need of any additional information about the user itself or the recommended entity. While such methods can be highly performant, they usually do not offer a straightforward explanation as to *why* a certain product or movie is being recommended. Still, as long as users receive interesting recommendations, one can assume that this is not a particular issue for the latter. Patients, in contrast, may be highly interested in solutions that not only fit their personal medical profile insofar, as they are built on medically meaningful information about the patient, but also provide explanations of the recommendation itself. That is to say, patients will arguably prefer recommendations optimised towards their individual medical needs, instead of suggestions based on the similarity to other patients that may show very similar behavioural patterns but have a completely different medical background. Thus, adding such a dimension to HRS may increase patients' trust into the model and ultimately improve patients' quality of care, as trust fosters not only the patient-doctor relationship, but also treatment adherence

(Fiscella et al., 2004). Meanwhile, from a business value perspective, healthcare providers offering medically meaningful recommendations can treat this as a value proposition to their clients thereby meeting current market trends of medical personalisation.

That being said, the resultant question is how to add such a medical personalisation and explanatory perspective to a ML model? A simple solution to this issue appears to be the incorporation of domain knowledge about patient's health conditions. In medicine, a viable candidate for such domain knowledge may for instance be the ICD-9 (International Classification of Diseases) hierarchy, since it is a comprehensive catalogue of all sorts of medical conditions published by the WHO and globally used among medical institutes. Yet, using data such as the ICD-9 hierarchy as a source of domain knowledge – independently of its evident benefits – faces a series of practical implementation problems. Chief among those is the structure of the data itself, as hierarchies are in nature tree-like structures hard to embed into the continuous space necessary for most ML models. Nevertheless, recent works of Nickel and Kiela (2017) and Chamberlain et al. (2017) proposing hyperbolic embeddings for learning hierarchical representations appear to provide a bypass for this issue, as will be elaborated further throughout this paper.

In light of the research problem outlined above, the objectives of this work are a) to investigate how to incorporate complex domain knowledge, such as the ICD-9 hierarchy into a HRS and b) examine whether said domain knowledge can add value to the model in terms of performance. In order to achieve this goal, this research proposes a content-based RS for patient-doctor matchmaking built on real data from a leading Portuguese healthcare provider. Thereby, patients' available ICD-9 records are used as a source of domain knowledge and enhanced with an external, pre-trained dataset of ICD-9 embeddings in the hyperbolic space.

More precisely, the following approach will be pursued: Contextualising the topic, chapter II begins with a bibliographical examination of related works laying out the general mechanics of RS and considering a series of practical applications of HRS. Additionally, in a rather theoretical part the concept of embedding hierarchical data into the hyperbolic space and its benefits will be examined, laying the foundation for the later proposed model. Notably, it will be shown why hyperbolic embeddings are inherently better equipped than their Euclidean counterparts to embed hierarchical, tree-like data into the continuous space. Moving forward, chapter III sheds light on the different data sources used to build this model, while pointing out the substantial data processing challenges connected to translation issues between different medical terminologies. Moreover, some exploratory data analysis will be performed in order to gain a better understanding of the data at hand. Approaching the core work of this paper, chapter IV is dedicated to the implementation of the model itself. After first introducing the notion of hyperbolic distance as a similarity measure for recommender systems, this section serves mainly for the formulation of two content-based models using hyperbolic distance, namely a patient ICD-9 similarity model and a doctor ICD-9 similarity model. For evaluation purposes, a conventional content-based model is formulated to serve as a benchmark. Ultimately, chapter V analyses the results of this investigation. As will be shown, incorporating complex domain knowledge to a HRS is not only feasible, but also adds value to the model. In fact, one of the proposed models, namely the doctor ICD-9 similarity model, outperforms its conventional counterpart in common evaluation metrics. With these findings in place, their implications for the healthcare sector will be discussed in the final chapter emphasising the potential business value of a HRS providing medical personalisation. Summarizing, this work understands itself as an initial investigation into the notion of using hyperbolic embeddings to incorporate domain knowledge into ML applications encouraging further research in this area. With that in mind, it has to be noted that the scope of this paper is limited to the proposed content-based models, while it is acknowledged that further research into hybrid solutions may yield in promising results.

II. Background and related work

To shed light on the broader context of this paper, this chapter provides a brief overview of recent literature on HRS as well as an outline of the functioning of RS in general. Furthermore, recent works on representation learning in the hyperbolic Poincaré space and their theoretical background will be discussed in order to understand how they might help incorporating complex domain knowledge into ML models.

A. Recommender systems in healthcare

Successful applications of ML can be found across different industries and domains ranging from spam and malware detection, over online chat-bots, stock price predictions and sentiment analysis, to product recommendations or social media services (McKinsey, 2018). Nevertheless, as prevalent as ML may be across sectors, within the healthcare domain, ML applications and particularly RS are only recently emerging, due to elevated requirements regarding reliability and trustworthiness, as well as increased data privacy regulations (Ramakrishnan et al, 2001; Özkan and Sezgin, 2013). The last years, however, have shown an increase in studies and research papers on HRS (Schäfer et al., 2017).

Among those works, medical user profiling and medical personalisation are two of the most popular topics. Their applications range from health and fitness apps, emphasising the user profiling aspect, to clinical decision support systems, focusing on medical personalisation. For instance, Luo et al. (2012) propose an application recommending relevant medical home care products based on a patient's individual history of health records. Going one step further, Radha et al. (2016) add a behavioural perspective to HRS in investigating the impact of different motivation and engagement strategies to improve lifestyle adaption recommendations for hypertension prevention and treatment. Furthermore, RS also find usage in therapy decision support. Gräßer et al. (2017) suggest incorporating inherent methods of

RS into clinical decision support systems in order to leverage the large volume of clinical data, instead of relying solely upon rule-based expert knowledge for decision support. Ultimately, Han et al. (2018) address the topic of patient-doctor matchmaking laying the foundation for this adjacent work. The authors propose a RS for suggesting primary care doctors to patients based on their prior consultation history and metadata.

In general, RS are a sub-class of information filtering systems with the goal to provide meaningful suggestions to users for certain items or entities, by attempting to predict the affinity or preference of a given user for said items (Melville and Sindhvani, 2010).[1] RS can be broadly divided into three major categories: collaborative filtering (CF) approaches, content-based (CB) recommenders and hybrid models, which are a combination of the former two. CF approaches rely solely upon past interactions recorded between users and items, whereas CB approaches use additional information about users and/or items (Ricci et al., 2011). More precisely, CF capitalises on behavioural data, i.e. users' co-occurrence patterns in order to detect similar users and/or items and make predictions based on these similarities, while CB recommenders explore user or item metadata to derive user preferences and model the observed user-item interactions.

As for CF approaches, a common challenge is the enormous size and the sparsity of the user-item interaction matrix, as most users naturally have only interacted with a small set of items. That being said, memory-based CF approaches, where one attempts to detect the nearest neighbour relying on the entirety of the large and sparse user-item interaction matrix can be computationally expensive and difficult to scale (Pennock et al., 2000). Consequently, model-based realisations of CF such as matrix factorisation (MF) are commonly used for this task, as

---

[1] Given that this work tackles the challenge of patient-doctor matchmaking, the terms "users" and "items" can be replaced by "patients" and "doctors". When speaking of RS in a broader context, however, the former appear more appropriate. Thus, they will be used interchangeably according to context for the remainder of this paper.

they are more robust against the aforementioned problems. The underlying rationale behind MF approaches is to decompose the huge and sparse interaction matrix into the inner product of two smaller and denser matrices, which are essentially latent feature representations of users and items, respectively, and letting the model learn these features (Koren et al., 2009).

In addition to sparsity, CF approaches often suffer from the cold-start problem, i.e. the question of what to do with new users that have no prior interactions usable for predictions (Ghazanfar and Prugel-Bennet, 2010). A way to bypass this issue is to use CB recommenders, since those do not depend on prior interaction data for predictions. Describing new items or new users based on their features instead, enables meaningful recommendations for new entities, even when there is no interaction data available. Despite the advantage with the cold-start problem, however, CF approaches tend to outperform CB recommenders, as usually even a few ratings are more valuable than metadata about users or items (Pilászy and Tikk, 2009). Ultimately, a method to balance both CF and CB respective limitations is to use hybrid recommenders, which are a combination of the former and the latter (Burke, 2002).

B. Hyperbolic embeddings

As has been hypothesized in the introduction, HRS might profit more than other areas from incorporating domain knowledge into the model. Within the healthcare context, such knowledge may include, for instance, a catalogue and categorisation of health conditions, such as the ICD-9 hierarchy. Abstracting a hierarchy into mathematical terms, it is essentially a complex tree that is defined as a connected graph in which for any pair of two vertices $u \neq v$ there is exactly one path connecting them (Bender and Williamson, 2010). An inherent characteristic of hierarchies or trees, however, is that they are discrete structures and thus embedding them in a way that can be used in machine learning models can be challenging, as the latter often rely on continuous representations (De Sa et al., 2018a). Thus, the underlying question is, how to efficiently and accurately model an increasingly complex hierarchy – and accordingly an increasingly complex tree – into a continuous space, such that the inherent information of the hierarchy can be used for machine learning?

Recent proposals of Nickel and Kiela (2017) and Chamberlain et al. (2017), suggesting hyperbolic embeddings for learning hierarchical representations to address this issue, have found much notice in the machine learning community. The rationale is that embeddings of hierarchical, tree-like data into the hyperbolic space perform better in the task of capturing and preserving the distances and complex relationships within a given hierarchy, than embeddings in the Euclidean space would. As a matter of fact, their works show that hyperbolic embeddings, even in very low dimensions, consistently outperform their higher-dimensional, Euclidean counterparts when learning hierarchical representations.

The reasons for said superiority lie within the properties of hyperbolic geometry itself. Hyperbolic space is a space with a constant negative curvature, in which disk area and circumference both grow exponentially with their radius – even in two dimensions. In fact, this very exponential expansion property renders it inherently well-suited for the task of embedding a tree into the continuous space, since for a regular tree the number of children grows exponentially with their distance to the root, as well (Nickel and Kiela, 2017). Meanwhile, Euclidean geometry would need to increase the dimensionality of the embedding, since in a 2-d Euclidean space circumference and disk area only increase linearly and quadratically with the radius, respectively.

While the considerations above hold for hyperbolic space in general, the preferred geometrical models for representation learning tasks, such as the one at hand, are the Poincaré disk for a 2-d hyperbolic geometry and the Poincaré ball model for 3-d or more. The main characteristic of Poincaré models is that all points of the geometry are interior of the unit disk (for 2-d) or the 3 or n-dimensional unit ball (for more than 2-d) (Anderson, 2005). Moreover, Poincaré models offer conform mapping between hyperbolic and Euclidean space, as angles are preserved, which is convenient when translating between spaces and models.

Recalling that the goal when embedding tree-like graphs into a continuous space is to preserve original graph distances, one needs to consider the hyperbolic distance, which is defined as $d_H(x,y) = \operatorname{acosh}\left(1 + 2\frac{||x-y||^2}{(1-||x||^2)(1-||y||^2)}\right)$ (1). In hyperbolic space, the shortest paths between two points, called geodesics, are curved (similarly to the space itself). Due to this curvature, the distance from the origin to a given point $d_H(O,x)$ grows towards infinity as x approaches the edge of the disc, as can be observed in figure 1.

Now, considering the embedding of a graph (or tree) into a continuous space, suppose x and y are children of a parent z, which is placed at the origin O. Then, the distance between x and y is $d(x,y) = d(x,O) + d(O,y)$ (2). Normalizing this equation, provides the distance ratio of the original graph, i.e. $\frac{d(x,y)}{d(x,O)+d(O,y)} = 1$ (3). This equation will be relevant in the following, since when comparing its behaviour in hyperbolic and Euclidean space, quite different effects can be observed. As is visualised in figure 2, when moving towards the edge of the unit disk, i.e. $x \to 1$, in Euclidean space $\frac{d_E(x,y)}{d_E(x,O)+d_E(O,y)}$ remains a constant, whereas in hyperbolic space $\frac{d_H(x,y)}{d_H(x,O)+d_H(O,y)}$ approximates 1, which is exactly the original graph distance ratio! Therefore, it can be seen that Poincaré embeddings are inherently better suited for this kind of representation learning task, due to their better capacity to preserve original graph distances with arbitrarily low distortion (De Sa et al. 2018a).

A further investigation of the detailed mathematics of Poincaré embeddings, as laid out by Nickel and Kiela (2017), Chamberlain et al. (2017) or De Sa et al. (2018b) amongst others, would be beyond the scope of this paper. Nevertheless, a brief overview of the actual applications of Poincaré embeddings seems appropriate.

First and foremost, De Sa et al. (2018b) provide a variety of powerful use-cases of representation learning, most of which related to natural language processing (NLP). The authors use their combinatorial construction algorithm to obtain entity embeddings of different types of hierarchies, such as WordNet's hypernym relationships or MusicBrainz' artist, album and song hierarchy. In light of the given topic, however, their work on embeddings of the UMLS diagnostic hierarchy from ICD-9 vocabularies is of particular interest. In fact, the provided ICD-9 embeddings contribute substantially to the proposed model and will therefore be described in more detail in chapter III.

Furthermore, Beaulieu-Jones et al. (2019) examine an equally tangible application of Poincaré embeddings to the healthcare domain, in general, and the ICD-9 hierarchy, in particular. In line with a core hypothesis of this paper, they argue that embeddings in the hyperbolic space could provide a better representation of diseases and medical concepts. Applying Poincaré embeddings to a large-scale dataset of electronic healthcare records, the authors show that hyperbolic embeddings of medical concepts outperform traditional approaches in terms of representations on real data. In addition, the paper illustrates that Poincaré embeddings can accurately reconstruct the ICD-9 hierarchy on synthetic data.

III. Data

While the theoretical benefits of Poincaré embeddings have been laid out in depth in the prior chapter, it remains to be examined whether they can add value in practice. For that purpose, a pre-trained dataset of Poincaré embeddings of the ICD-9 hierarchy has been applied to real data from a leading private clinical network in Portugal. Thus, the following chapter will provide an overview of the data sources used, briefly explain the necessary processing and mapping steps between them, and ultimately perform an exploratory data analysis of the final dataset.

A. Data sources

By and large, four different data sources have been used: 1) data from the healthcare service provider itself, 2) the ICD-9 hierarchy master data, 3) Poincaré embeddings of the ICD-9 hierarchy in Unified Medical Language System (UMLS) entities provided by Hazy Research and 4) mapping keys between UMLS, SNOMED CT[2] and ICD-9 terminologies.

The data provided by the Portuguese clinical network is the core data of the project. The private healthcare provider operates country-wide with 18 hospitals or clinics across Portugal. According to its website, it serves more than 1,000,000 patients p.a. with over 2,000,000 consultations and 8,000,000 examinations and has more than 5,000 members of staff. Building on the prior work of Han et al. (2018), an already curated dataset has been used for this paper. Said data can be divided into three categories: First of all, there are demographic information, such as gender, age or location of the patients themselves. For inpatients, i.e. patients who stay at the hospital while under treatment, this demographic information is enriched with medical information in form of the ICD-9 code. Analogously, the dataset contains demographic and professional information, such as gender or the hospital of the doctors. Ultimately, the interactions between patients and doctors are captured as episodes.[3]

The aforementioned medical information about the patients are based on ICD-9 codes, which are the global standard for reporting diseases and health conditions published and maintained by the WHO. As such, the ICD is the endeavour to comprehensively catalogue, classify and map all sorts of "diseases, disorders, injuries or other health conditions" (WHO, 2020a). In practical terms, the ICD is a hierarchy of codes, each of which corresponding to a specific health condition. Since this system of classifications evolved over time, there are different

---

[2] Systematized NOmenclature of MEDicine Clinical Terms

[3] In medical jargon, interactions between patients and doctors are referred to as "episodes of care".

ICD versions in place. As of the date of the publication of this paper, the newest version is ICD-11, which was published in 2018 and will become effective for reporting in 2022.

Given that the underlying investigative question of this paper is whether complex domain knowledge, such as the ICD-9 hierarchy embedded into hyperbolic space can add value to HRS, the need for such embeddings in the first place is evident. While it would be possible to derive embeddings directly from the healthcare provider dataset by for instance analysing co-occurrence patterns among diseases, it can be argued that such an undertaking would show little promise due to the relatively small size of the dataset, as well as the fact that not all ICDs occur in it, which might distort graph distances. Instead, this work aims to capitalise on a pre-trained dataset of Poincaré embeddings provided by Hazy Research. The research group used the diagnostic hierarchy of ICD-9 vocabulary in the UMLS to retrieve Poincaré embeddings of single medical concepts within the ICD-9 hierarchy. This method results in unique hyperbolic embeddings of medical concepts (identified by the CUI, i.e. Concept Unique Identifier) available in different levels of dimensionality (10, 20, 50 or 100d). The model developed in this paper bases on the 100d embeddings. As such, the embeddings classify as transfer learning data subjecting them to a set of limitations including the risk of negative transfer or overfitting (Torrey and Shavlik, 2010), in addition to general mapping problems. Nonetheless, it needs to be stressed that the pre-trained embeddings are paramount to the success of this research, as they represent the actually meaningful information aimed to be incorporated into the proposed model.

Ultimately, since the given hyperbolic embeddings are only available in UMLS and not directly available for ICD-9 codes, a mapping between UMLS and ICD-9 is needed. As will be explained further in the following chapter, this process is by no means trivial, due to the fact that there is no direct mapping from CUI codes in UMLS to the ICD-9 codes of the core dataset. Consequently, the SNOMED CT were selected as an intermediary, as it is a very comprehensive healthcare terminology and transferable both to UMLS' CUIs and ICD-9 codes (SNOMED International, 2020). All UMLS data including SNOMED CT and CUIs have been retrieved from the US National Library of Medicine (NLM).

B. Data processing

In light of the considerations above, mapping the hyperbolic entity embeddings in UMLS' CUIs to the mere ICD-9 codes of the core dataset represents a substantial challenge during the data processing stage. A mapping between SNOMED CT and ICD-9 is updated and published annually by the US NLM and could be added to the ICD-9 data straightforwardly (US NLM, 2019). In order to retrieve a mapping between CUI and SNOMED CT, however, the multilingual medical terminology service PyMedTermino has been used. PyMedTermino is an open source API that provides access to the main medical terminologies, as well as mappings between those terminologies (Lamy, 2015). In practice, the Python implementation of PyMedTermino was used to parse through the UMLS Metathesaurus (provided by the US NLM) to access the desired CUIs and subsequently map them to the SNOMED CT.

Figure 3 provides an overview of the data curation process. There are two main flows of data: the core dataset from the healthcare provider and the entity embeddings enriched with corresponding terminology mappings. Regarding the entity embeddings part, at first, the CUI to SNOMED CT and the SNOMED CT to ICD-9 mapping data was merged on the SNOMED CT in order to obtain a CUI to ICD-9 mapping. This enabled a merge of ICD-9 codes with the hyperbolic entity embeddings. Having established a unique hyperbolic embedding per ICD-9 code, these embeddings needed to be added to the core dataset. Naturally, only the patients having an ICD-9 record are included in the final dataset. These patients and their ICD-9 codes were enriched with the hyperbolic embeddings, such that only a list of patients that have at least one ICD-9 diagnosis, which in turn has an entity embedding, remain. Consequently, all episodes with non-ICD-9 patients were excluded, as well as doctors who only had episodes with such patients.

At this point, it needs to be acknowledged that this data curation process suffers from two main limitations, namely the lack of mapping and, if there is a mapping, the quality of such (compare table 1). As ML models often capitalise on huge amounts of data, it is unfortunate that a lot of valuable information is lost due to missing translations between medical terminologies impeding transfer learning tasks significantly. Moreover, as mappings between SNOMED CT, ICD and CUI terminologies are being improved only incrementally, they can by no means be considered flawless. Hence, the potentially meaningful contribution of pre-trained domain information in form of ICD embeddings is reduced substantially, if applied to the wrong diseases, as ultimately a model can only be as good as the data that feeds it.

C. Exploratory data analysis

Despite the substantial reduction from its initial size, a dataset consisting of over 33k patients and 223 doctors with more than 166k interactions between them remain. Thus, in the following a variety of insights relevant to the patient-doctor matchmaking process will be laid out and analysed.

To begin with, figure 4 illustrates an increasing demand for primary care doctors with roughly 27k episodes in 2012 and 34k in 2016, underlining the relevance of primary care and consequently ML applications in this domain.[4] Furthermore, as it can be pertinent how often patients switch doctors, or stick with the same, it is worthwhile analysing the frequencies of visits as well as the “churn” rate of patients. Notably, over the observed period patients only had a total of five visits and saw 2.6 distinct doctors on average. More precisely, almost 65%

---

[4] Note that for 2017 only the first four months are available.

of patients visited at least two different doctors, whereas only 28% visited the same doctor twice. While this indicates that patients tend to switch their family doctors quite frequently not trusting their family doctors right away, it has to be noted that this dataset only contains inpatients, who arguably are more likely to have more health records, visit doctors more frequently and, hence, have a higher likelihood to have seen multiple doctors. Doctors, on the other hand, had 745 episodes on average over the observed period, with an average of 381 distinct patients. Considering patient demographics, it is little surprising that a large portion of patients are elderly people or at least adults (compare figure 5). The average patient age is 55. Finally, the gender distribution both among patients and doctors is relatively balanced, albeit with a slight surplus of female doctors.

With this work's strong emphasis on the patient's ICD-9 information, it is of relevance to analyse how the different diseases distribute in the given dataset. The ICD-9 hierarchy is structured by 18 overarching chapters, which in themselves are divided into sub-chapters. These sub-chapters, then, contain major ICD codes, which in most cases have further specifications of a particular medical condition. As figure 6 displays, the most common ICD-9 chapters are injuries and poisoning, followed by diseases of the circulatory system. Both observations are intuitive, as naturally many patients consult their doctors with effects from injuries, and circulatory diseases are very common throughout the population in general and among elderly in particular. This is underpinned further, when analysing the levels below the two most common chapters, as figures 7 and 8 illustrate. For instance, in the injuries and poisoning chapter, the most common sub-chapters are the late effects of injuries or poisoning themselves, as well as "other and unspecified effects of external causes" including effects of heat and light. All of which are arguably very common causes to visit a primary care doctor. The same is the case for diseases of the circulatory system, where hypertension – also a fairly common disease – is the largest sub-chapter.

Ultimately, to complete the exploration of the data used in this work, a few considerations on the hyperbolic embeddings are adequate. Due to the high dimensionality of the data, a visualisation of the embeddings is not possible. Nevertheless, analysing hyperbolic distance measures between single ICD-9 codes within the dataset shows that inherent characteristics of the hierarchy have been captured well by the embeddings, as the distances between related diseases, i.e. from the same chapter or sub-chapter, tend to be small, whereas distances between unrelated ICD-9 codes tend to be bigger. When reducing the dimensionality to 2-d, however, one can, indeed, illustrate how a hierarchy embedded into a graph looks like, as the Poincaré disk model's properties allow for a translation into Euclidean geometry. For instance, the Poincaré embeddings of the ICD-9 hierarchy built on synthetic data in two dimensions by Beaulieu-Jones et al. (2019) exemplify this quite well. Figure 9 shows how single ICD-9 codes are clustered around their respective chapters and sub-chapters. Whilst acknowledging the limited comparability of the separate embedding datasets, due to the different data and methodology used, it can be observed, how well Poincaré embeddings, in principle, capture the ICD-9 hierarchy in terms of preserving original graph distances.

IV. Methods

With the theoretical foundation set and the properties of the data analysed in depth, a model using Poincaré embeddings can be formulated. For that purpose, the following chapter first introduces the notion of using hyperbolic distance as the similarity measure in RS explaining its differences to conventional approaches and, subsequently, examines the mathematical implementation of this approach in practice.

A. *Hyperbolic distance as a similarity measure for recommender systems*

As has been elaborated in chapter II, the goal of RS is to predict the affinity or preference of users for certain items or in this case patients and doctors. This is typically achieved by, first, deriving a similarity measure of all users or all items and, subsequently, use this similarity measure as a weight in combination with past interaction data to predict users' preference for a given item. Despite this basic concept, heuristics to retrieve said similarity measure vary in the type of data used as features, as well as the selected mathematical method. For instance, collaborative filtering approaches rely on past user-item interaction data, whereas content-based methods use metadata. Different mathematical methods for similarity detection, on the other hand, include cosine similarity, Pearson correlation, Jaccard similarity or mean-squared differences among others (Agarwal and Chauhan, 2017).

The most popular methods within the context of RS, however, are cosine similarity and Pearson correlation – in the following referred to as s_c and s_P, respectively. s_c measures the similarity of two vectors by considering the cosine of the angle between them. As such, for two vectors A and B, it is defined as $\cos(\theta) = \frac{A*B}{||A||*||B||}$ (4). s_c assumes values between 0 and 1, with $\cos(0°) = 1$ being the highest similarity score. Treating users' preference vectors as variables, s_P, on the other hand, measures the linear correlation between two given users (or items). For the variables A and B, it is defined as $\rho(A, B) = \frac{cov(A,B)}{\sigma A \sigma B}$ (5). s_P assumes values between -1 and 1, with -1 indicating perfect negative correlation, 1 indicating perfect positive correlation and 0 indicating no correlation at all. In terms of similarity, the closer to 1 the correlation converges, the higher the similarity between users or items.

Besides their simplicity and easy interpretation, both s_c and s_P are typically preferred over simple Euclidean distance, since – in case of explicit feedback – they are robust against individually different rating scales among users. For instance, two users may have a very similar taste, but different rating generosity with one being rather critical tending to rate items below average, whereas the other typically rates items comparatively higher. Considering their Euclidean distance, they may appear quite different from one another. Nonetheless, both

s_C and s_P are able to capture this *hidden* similarity, since s_C is ultimately a measure of angle and not magnitude and s_P captures linear trends by design.

At this point it must be noted, however, that both s_C and s_P (as well as most other conventional similarity measures in RS) are inherently suited for the Euclidean space only (Leimeister and Wilson, 2019). That being said, for the purpose of this paper the notion of hyperbolic distance as a more suitable similarity measure needs to be evaluated. As each of the derived ICD-9 diseases is represented by an embedding in the hyperbolic space, the objective is to determine how similar these diseases – and ultimately the patients admitted with or doctors having treated these diseases – are with one another.

The basic principle of hyperbolic distance as a similarity measure is simple: Once a unique embedding per either patient or doctor is derived, a patient-patient or doctor-doctor similarity score can be determined for each patient or doctor by utilising the hyperbolic distance function from equation (1). The resulting matrix of distances is subsequently scaled from 0 to 1 and subtracted from 1 in order for 0 to be the minimum similarity and 1 the maximum. Applying this heuristic, yields a similarity score that is not only consistent with hyperbolic space (i.e. preserving the hierarchal information and complexities of the input graph), but also as easily interpretable as s_C or s_P. This similarity measure shall be referred to as *hyperbolic similarity* s_H for the remainder of this paper and its implementation into the model at hand will be examined in the following chapter.

B. Implementation of a content-based recommender system using hyperbolic distance

Since the ICD-9 embeddings represent metadata about patients or doctors, a RS using said embeddings classifies as content-based. While there is further data available for both patients and doctors (e.g. demographic or location data), the proposed model will consider only the

ICD-9 information. In fact, it will be benchmarked against an alternative content-based model using that very metadata for performance evaluation purposes in the next chapter.

As has been discussed in chapter III, the ICD-9 information per patient from the core dataset has been enriched with the Poincaré embeddings provided by Hazy Research. Since many patients have been admitted with more than one disease throughout their individual medical history, naturally, the majority of patients have multiple ICD-9 entries. Therefore, in order to determine a unique embedding per patient and per doctor, multiple entries need to be averaged – especially in the case of doctors who obviously treat many different patients. Due to the specific properties of the hyperbolic space, however, the usual Euclidean mean is not applicable and thus a generalisation is needed. In hyperbolic geometry, the averaging of feature vectors is done by using the Einstein midpoint (Khrulkov et al., 2019). The Einstein midpoint takes its simplest form in Klein coordinates and is defined as $HypAve(x_1, \ldots, x_N) = \frac{\sum_{i=1}^{N} \gamma_i x_i}{\sum_{i=1}^{N} \gamma_i}$ (6), where $\gamma_i = \frac{1}{\sqrt{1-c||x_i||^2}}$ and *c = 1* as *c* corresponds to the radius of the Poincaré disk. The Klein model is consistent with the Poincaré ball model[5], but since the same point has different representations in the two models, they need to be first translated from the Poincaré to the Klein model, then averaged and ultimately mapped back into the Poincaré model in order to complete the operation. Thus, if $x_{\mathbb{D}}$ and $x_{\mathbb{K}}$ correspond to the same point in the Poincaré and the Klein model, respectively, then the following formulas serve for translating between them: $x_{\mathbb{D}} = \frac{x_{\mathbb{K}}}{1+\sqrt{1-c||x_{\mathbb{K}}||^2}}$ and $x_{\mathbb{K}} = \frac{2x_{\mathbb{D}}}{1+c||x_{\mathbb{D}}||^2}$.

---

[5] In contrast to the Poincaré model, the Klein model considers lines as straight Euclidean lines, whereas in the former lines are arcs that are orthogonal to the boundary. Moreover, the Klein model is not conformal, meaning that angles and circles are not preserved in the curvature of the space.

With an appropriate methodology for hyperbolic feature vector averaging in place, a content-based model for patient-doctor matchmaking can be formulated. In formal terms, for N patients and K doctors, the patient-doctor interaction matrix $\mathrm{Y} \in \mathbb{R}^{\mathrm{N} \times \mathrm{K}}$ is denoted as:

$$y_{ij} = \begin{cases} 1, & if\ patient\ i\ interacted\ with\ doctor\ j \\ 0, & otherwise \end{cases}.$$

Adapting the work of Han et al. (2018), the patient-doctor interactions are furthermore weighted with a trust measure. Thereby, the trust between a patient and a doctor is modelled in dependence of the recency and frequency of their consultation history, i.e. doctors that have been visited repeatedly and recently will be weighted higher for a given patient.

Regarding feature creation, the ICD-9 embeddings need to be considered. If $\mathrm{V} \in \mathbb{R}$ is the set of all Poincaré embeddings, with each embedding being essentially a 1×100 dimensional row vector, then for each patient i the set of embedding vectors is denoted as $\mathrm{V}_i \subset \mathrm{V}$ corresponding to all ICDs that patient has been diagnosed with. Similarly, for each doctor j the set of embedding vectors is specified by $\mathrm{V}_j \subset \mathrm{V}$ corresponding to the ICDs of all patients that visited said doctor. Hence, the feature vectors of patient i and doctor j are given by the hyperbolic average of their embeddings, i.e. $f_i = HypAve(V_i)$ and $f_j = HypAve\left(V_j\right)$ (7).

With the feature matrices for patients and doctors established, the similarity across patients and doctors can be calculated. For purposes of simplicity, this process will be described only for doctor-doctor similarity, while it is acknowledged that the method is analogously applicable for patients. The similarity between doctor j and k is described by the above defined *hyperbolic similarity* of their feature embeddings, i.e. $s_{j,k} = s_H(f_j, f_k)$ (8).

Ultimately, the predicted affinity $p_{i,j}$ of a user i towards a doctor j can be computed using the following operation: $p_{i,j} = \frac{\sum_{k=1}^{K} y_{i,k} * s_{j,k}}{\sum_{k=1}^{K} s_{j,k}}$ (9). Recalling that K is equal to the total amount of doctors and $y_{i,k}$ is the trust-weighted interaction value between patient i and doctor k, it becomes evident that the predicted affinity of patient i is essentially given by the similarity-weighted sum of doctors the patient visited previously, divided by the sum of the weights. While RS in e-commerce and such usually aim to suggest primarily new, unseen items, this model does not exclude doctors the patient already interacted with for recommendation. This is of relevance insofar, as the goal of this model is to suggest the patient with the best suiting doctor for their next primary care visit, for which previously seen doctors are arguably highly relevant candidates and should by no means be excluded.

V. Results and discussion

While a substantial part of this paper has been dedicated to the theoretical benefits of Poincaré embeddings and their application to the given problem of patient-doctor matchmaking, it is ultimately necessary to evaluate their performance in comparison to conventional methods, in order to judge their actual value. Therefore, this chapter will evaluate the proposed models with respect to common performance metrices and furthermore discuss their value from a commercial perspective.

A. Proposed models

In line with the methods described above, this paper proposes the following models:

1. **Conventional CB:** a patient-patient-similarity based benchmark model using s_c of patient features such as gender, age, location, as well as 1-hot encoded ICD-9 data to capitalise on co-visitation patterns of patients with similar demographic profiles,
2. **Patient ICD-9 similarity**: a patient-patient-similarity based recommender capitalising on co-visitation patterns of patients with similar diseases and,
3. **Doctor ICD-9 similarity**: a doctor-doctor-similarity based recommender aiming to identify doctors that have similar expertise to the ones the patient visited in the past.

For each model a list of either 3, 5 or 10 suggested doctors is generated, which hereinafter will be analysed with respect to their performance.

B. Model evaluation

For the performance evaluation of RS, it is common practice to rely on overall root mean square error (RMSE) or hit rate (HR) and precision (p) as evaluation criteria (Chen and Liu, 2017). However, since most RS – including this one – are typically presented as a sorted list, the latter are arguably more suitable, as the evaluation objective is to see, if the patient actually visited *one* of the recommended doctors or not. That being said, *HR@n* refers to the number of total hits, divided by number of patients depending on the number of recommended doctors *n*. Analogously, *p@n* indicates the amount of correctly predicted doctors depending on *n*. Intuitively, the HR will increase with a growing number of recommendations, whereas p will decrease. As a matter of fact, the very reason to combine these two evaluation criteria is that although it is desirable to maximise the number of hits, patients should not be confused with too many options that do not meet their needs, as this might even have counterproductive effects, e.g. paradox of choice (Schwartz, 2004).

Figure 10 illustrates the performance of the three suggested models regarding HR and p. While the patient ICD-similarity model is apparently not able to add substantial value scoring even slightly below the benchmark model, the doctor ICD-similarity model does, indeed, outperform the benchmark model. In fact, this allows for two major conclusions in light of the theoretical considerations in the chapters above: First, hyperbolic averaging appears to be a viable method for feature averaging of Poincaré embeddings considering the substantial number of different patients and diseases doctors treat. This is insofar noteworthy, as one might reasonable assume that the more disease embeddings are being averaged, the less meaningful they become. Yet, the resulting averaged embeddings are evidently still capable to set apart doctors fairly well. And second, the Poincaré embeddings – despite not having been trained on this dataset – can add value to this HRS.

Returning to the underlying investigation question of this paper, these two observations show that Poincaré entity embeddings of hierarchical data are, indeed, a powerful framework to help incorporate complex domain knowledge into a ML application in the healthcare sector. With this in mind, the business implications for the healthcare sector remain to be considered.

C. Implications for healthcare sector

As has been hypothesised in the introduction, the potential business value of successfully incorporating complex domain knowledge into machine learning applications in the healthcare sector may be substantial. Recalling that the goal of a matchmaking algorithm between patients and doctors is insofar different from typical e-commerce RS, as it aims to recommend patients with the doctor best suited for their specific, medical condition, instead of the "next best doctor", different evaluation criteria may apply from a business value perspective. For instance, one might argue that technical performance evaluation metrices such as hit rate and precision are, in fact, negligible in favour of a more qualitative evaluation. RS, in general, often suffer from popularity bias, in that they tend to suggest mostly popular doctors (Abdollahpouri et al., 2019). That being said, patients should not be matched with doctors because they are popular or because other patients with similar demographics visited them (even if this yields in high HR and p scores), but because they best fit their medical needs. Hence, it is suggested for further research that recommenders akin to this work should be optimised not only with respect to hit rate and precision, since this may not fully account for popularity bias, but also towards the domain-specific quality of the recommendation.

In light of these considerations, healthcare providers can treat this factor as a value proposition for their clients. With increasing demand for personalised healthcare solutions, recommenders built on patients' individual health record are arguably in-line with current market trends. Picturing a potential customer journey, the RS would suggest a patient that has been admitted with, for example, hypertension with doctors that have treated many cases of hypertension or similar diseases. In addition, making recommendations based on individual health profiles adds an explanatory perspective to the suggestions that many RS lack. Since health is a sensitive topic, in general, and trust into AI solutions is a major concern in the healthcare domain, in specific, this may be a substantial driver for the success and adaptation of HRS in practice.

VI. Conclusion

Recapitulating the objectives of this work, the previous chapters have shown that incorporating complex domain knowledge in form of Poincaré embeddings of the ICD-9 hierarchy into a HRS is not only possible, but also yields in an actual performance improvement in comparison to conventional approaches. Based on prior research, this paper has examined the benefits of the hyperbolic space for representation learning tasks in theory and, furthermore, applied the concept to real dataset. In doing so, it has been shown that Poincaré embeddings can contribute meaningful value in domains going beyond their original scope of NLP. Ultimately, it has been argued that the incorporation of domain knowledge is of particular value in the healthcare domain, as it allows for medically personalised recommendations that are also insofar more transparent, as they base on the patient's individual health record.

While the results of this preliminary investigation in this field are promising in principle, a set of limitations remains to be considered. As a matter of fact, from each of said limitations direct recommendations for further research can be derived. First of all, since the proposed models are purely CB in nature, they neglect valuable information that can be retrieved from patient-doctor interaction data. That being said, further research on a hybrid RS leveraging both interaction data and ICD-9 embeddings is strongly encouraged. Indeed, despite being beyond the scope of this paper, preliminary experiments with such hybrid models have shown first, promising results, albeit still needing optimisation.

In addition, data quality and data consistency have been persistent issues throughout this work with a substantial portion of available data lost due to insufficient mapping between terminologies. As has been stressed before, transferability between terminologies is paramount to the further growth of AI in healthcare and healthcare analytics. Hence, the healthcare industry as a whole is strongly encouraged to foster collaboration among different initiatives such as SNOMED CT, UMLS, ICD and others.

Analogously to the need of improved data quality in the healthcare sector in general, healthcare service providers, in specific, need to drive the digitisation in their industry if AI is to play a role in future healthcare. For instance, instead of capturing only the ICD information of inpatients, all patients attended by a doctor should be assigned with a diagnostic code in order to increase the scalability of ML solutions as the one at hand. Additionally, since catalogues as the ICD evolve over time, healthcare institutions need to ensure their systems and processes are not outdated. Anecdotal evidence to this issue is the fact that as of the date of the publication of this paper the WHO only issued emergency COVID-19 codes for ICD-10, while an ICD-9 code is strikingly absent from the list (WHO, 2020b).

At last, a less technical and more content-related issue is the fact that primary care physicians tend to be generalists arguably diminishing the supposed benefit of medical specialisation of the recommendations. With primary care doctors treating a variety of very different patients throughout their consultation history, it is harder to determine a sharp, distinguishable profile for each doctor. Thus, repeating similar investigations in more specialised fields of healthcare than primary care may yield promising results in the future.

References

Pennock, David M., Eric Horvitz, Steve Lawrence and C. L. Giles. 2000. "Collaborative Filtering by Personality Diagnosis: A Hybrid Memory- and Model-Based Approach." *Uncertainty in Artificial Intelligence Proceedings 2000*: 473–80.

Ramakrishnan, Naren, Benjamin J. Keller, Batul J. Mirza, Ananth Y. Grama and George Karypis. 2001. "Privacy risks in recommender systems." *IEEE Internet Computing*, 5(4): 54–62.

Burke, Robin. 2002. "Hybrid Recommender Systems: Survey and Experiments." *User Modeling and User-Adapted Interaction*, 12(4): 331–70.

Fiscella, Kevin, Sean Meldrum, Peter Franks, Peter Shields, Paul Duberstein, Susan McDaniel and Ronald Epstein. 2004. "Patient trust: is it related to patient-centered behavior of primary care physicians?" *Med Care*, 42(11): 1049–55.

Schwartz, Barry. 2004. *The Paradox of Choice*. New York: Ecco.

Anderson, James. 2005. *Hyperbolic Geometry*. 2nd ed. London: Springer-Verlag London.

Bergman, Lawrence. 2009. *Proceedings of the third ACM conference on Recommender systems*. New York, NY: ACM.

Koren, Yehuda, Robert Bell and Chris Volinsky. 2009. "Matrix Factorization Techniques for Recommender Systems." *Computer*, 18: 42–49.

Pilászy, István and Domonkos Tikk. 2009. "Recommending New Movies: Even a Few Ratings Are More Valuable Than Metadata." *RecSys '09: Proceedings of the third ACM conference on Recommender systems*: 93–100.

Bender, Edward A. and S. G. Williamson. 2010. *Lists, Decisions and Graphs. With an Introduction to Probability*: Edward A. Bender & S. Gill Williamson.

Ghazanfar, Mustansar A. and Adam Prugel-Bennett. 2010. "A Scalable, Accurate Recommender System." University of Southampton.

Melville, Prem and Vikas Sindhvani. 2010. “Recommender Systems.” In *Encyclopedia of machine learning*, ed. Claude Sammut, and Geoffrey I. Webb, 1–8. New York, London: Springer.

Sammut, Claude and Geoffrey I. Webb, ed. 2010. *Encyclopedia of machine learning*. New York, London: Springer.

Torrey, Lisa and Jude Shavlik. 2010. “Transfer Learning.” University of Wisconsin.

Ricci, Francesco, Lior Rokach and Bracha Shapira. 2011. “Introduction to Recommender Systems Handbook.” In *Recommender Systems Handbook*, ed. Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor, 1–35. Boston, MA: Springer US.

Luo, Gang, Selena Thomas, and Chunqianq Tang. 2012. “Automatic Home Medical Product Recommendation.” *Journal of medical systems*, 36: 383-398.

Sezgin, Emre and Sevgi Ozkan. 2013. “A systematic literature review on health recommender systems.” *E-Health and Bioengineering Conference (EHB), 2013*: 1–4. http://ieeexplore.ieee.org/servlet/opac?punumber=6695804.

Lamy, Jean-Baptiste, Alain Venot and Catherine Duclos. 2015. “PyMedTermino: an open-source generic API for advanced terminology services.” European Federation for Medical Informatics. Studies in health technology and informatics.

Radha, Mustafa, Martijn C. Willemsen, Mark Boerhof and Wijnand A. IJsselsteijn. 2016. “Lifestyle Recommendations for Hypertension through Rasch-based Feasibility Modeling.” In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization - UMAP '16*, ed. Julita Vassileva, James Blustein, Lora Aroyo, and Sidney D'Mello, 239–47. New York, New York, USA: ACM Press.

Vassileva, Julita, James Blustein, Lora Aroyo and Sidney D'Mello, ed. 2016. *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization - UMAP '16*. New York, New York, USA: ACM Press.

Agarwal, Ajay and Minakshi Chauhan. 2017. “Similarity Measures used in Recommender Systems: A Study.” *International Journal of Engineering Technology Science and Research*, 4(6): 619–26.

Chamberlain, Benjamin P., James Clough and Marc P. Deisenroth. 2017. “Neural Embeddings of Graphs in Hyperbolic Space.” Imperial College London.

Chen, Mingang and Pan Liu. 2017. “Performance Evaluation of Recommender Systems.” *International Journal of Performability Engineering*, 13(8): 1246–56.

Gräßer, Felix, Hagen Malberg, Sebastian Zaunseder, Stefanie Beckert, Denise Küster, Jochen Schmitt and Susanne Abraham. 2017. “Neighborhood-based Collaborative Filtering for Therapy Decision Support.” *Proceedings of the Second International Workshop on Health Recommender Systems colocated with ACM RecSys 2017*: 1–5.

Nickel, Maximilian and Douwe Kiela. 2017. “Poincaré Embeddings for Learning Hierarchical Representations.” Facebook AI Research.

Schäfer, Hanna, Santiago Hors-Fraile, Raghav P. Karumur, André Valdez, Alan Said, and Helma Torkamaan. 2017. “Towards Health (Aware) Recommender Systems.” Associtation for Computing Machinery.

Ghassemi, Marzyeh, Tristan Naumann, Peter Schulam, Andrew Beam and Rajesh Ranganath. 2018. “Opportunities in Machine Learning for healthcare.” Massachusetts Institute of Technology; John Hopkins University; Harvard Medical School; New York University.

Han, Qiwei, Mengxin Ji, Martinez de Rituerto de Troya, Inigo, Manas Gaur and Leid Zejnilovic. 2018. “A Hybrid Recommender System for Patient-Doctor Matchmaking in Primary Care.” *2018 IEEE 5th International Conference on Data Science and Advanced Analytics*: 481–90.

McKinsey Global Institute. 2018. “Notes from the AI Frontier - Insights from Hundreds of Use Cases.” McKinsey & Company.

De Sa, Christopher, Albert Gu, Christopher Ré and Frederic Sala (a). 2018. “Hyperbolic Embeddings with a Hopefully Right Amount of Hyperbole.” Stanford University. https://dawn.cs.stanford.edu/2018/03/19/hyperbolics/.

De Sa, Christopher, Albert Gu, Christopher Ré and Frederic Sala (b). 2018. “Representation Tradeoffs for Hyperbolic Embeddings.” Stanford University; Cornell University. *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*: IEEE.

Abdollahpouri, Himan, Robin Burke and Bamshad Mobasher. 2019. “Managing Popularity Bias in Recommender Systems with Personalized Re-Ranking.” *The Thirty-Second International Florida Artificial Intelligence Research Society Conference (FLAIRS-32)*: 413–418.

Beaulieu-Jones, Brett K., Isaac Kohane and Andrew Beam. 2019. “Learning Contextual Hierarchical Structure of Medical Concepts with Poincairé Embeddings to Clarify Phenotypes.” *Pacific Symposium on Biocomputing 2019*: 8–17.

Khrulkov, Valentin, Leyla Mirvakhabova, Evgeniya Ustinova, Ivan Oseledets and Victor Lempitsky. 2019. “Hyperbolic Image Embeddings.” Skolkovo Institute of Science and Technology; Institute of Numerical Mathematics of the Russian Academy of Sciences; Samsung AI Center.

Leimeister, Matthias and Benjamin J. Wilson. 2019. “Skip-gram word embeddings in hyperbolic space.” Lateral GmbH.

US National Library of Medicine. 2019. ICD-9-CM Codes to SNOMED CT Map. [Online]. https://www.nlm.nih.gov/research/umls/mapping_projects/icd9cm_to_snomedct.html.

SNOMED International. 2020. SNOMED CT - 5 step briefing. [Online]. nomhttp://www.snomed.org/snomed-ct/five-step-briefing.

World Health Organization (a). 2020. International Classification of Diseases. [Online]. https://www.who.int/classifications/icd/en/.

World Health Organization (b). 2020. "International Guidelines for Certification and Classification (Coding) of COVID-19 as cause of Death." [Online]. https://www.who.int/classifications/icd/Guidelines_Cause_of_Death_COVID-19-20200420-EN.pdf?ua=1.

List of Abbreviations

Abbreviation	Definition
AI	Artificial Intelligence
API	Application Programming Interface
CB	Content-based
CF	Collaborative Filtering
CUI	Concept Unique Identifier
HR	Hit Rate
HRS	Healthcare Recommender System(s)
ICD	International Classification of Diseases
MF	Matrix Factorisation
ML	Machine Learning
NLP	Natural Language Processing
P	Precision
RMSE	Root Mean Square Error
RS	Recommender System(s)
SNOMED CT	Systematized NOmenclature of MEDicine Clinical Terms
UMLS	Unified Medical Language System
US NLM	US National Library of Medicine
WHO	World Health Organization

Appendix

A. Figures

Figure 1[6]: Poincaré disk model

Distance between points grow towards infinity when approaching the edge of the disk in the Poincaré disk model.

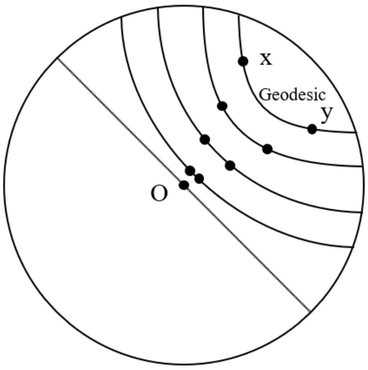
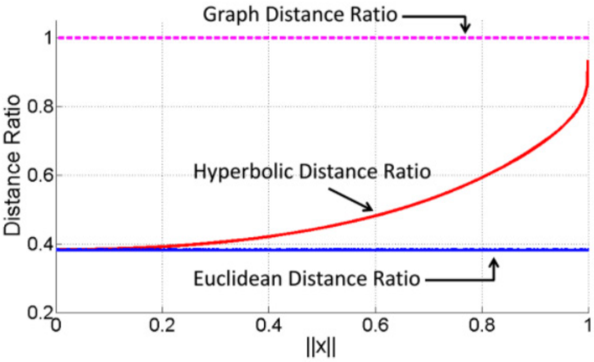


Figure 2: Distance ratios

Distance ratios of hyperbolic and Euclidean distance in comparison with original input graph distance ratio (De Sa et al., 2018a).



---

[6] Follow link on figure number to jump back to position in text.

Figure 3: Data diagram

Diagram describing the data sources, as well as the necessary mapping steps between terminologies and datasets.

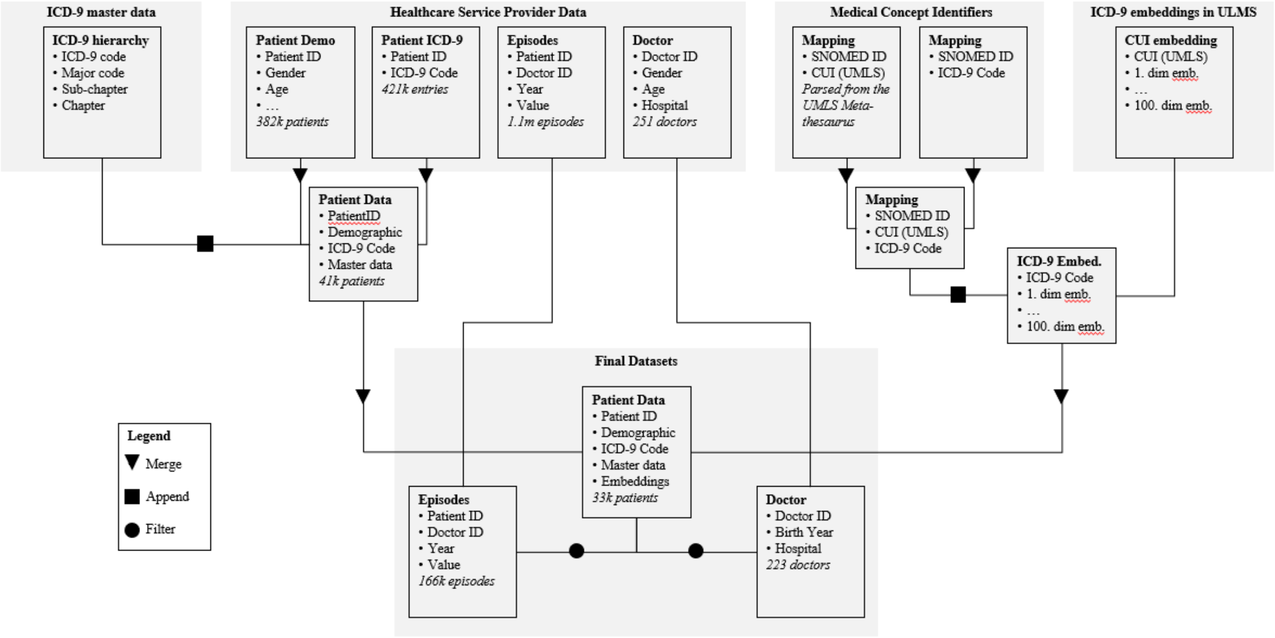


Figure 4: Number of episodes over time

Patient visits increased steadily over the observed period.

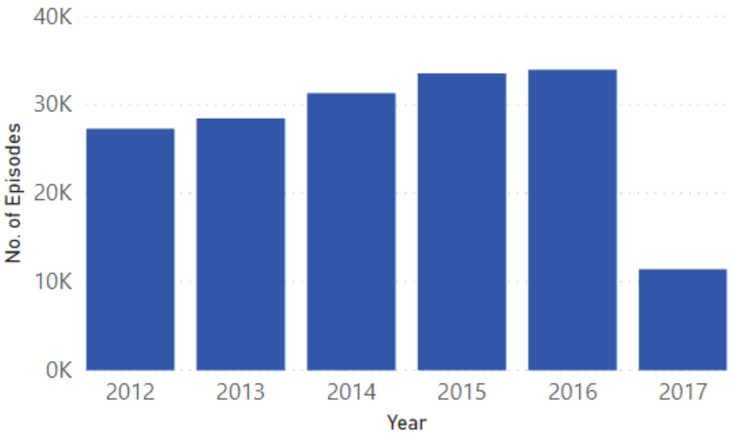


Figure 5: Patient age histogram

As can be observed, the majority of patients are either adults or elderly people.

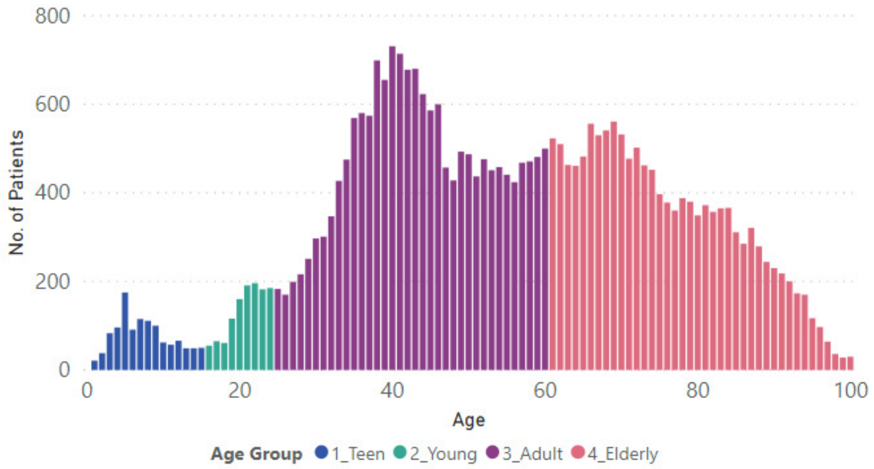


Figure 6: Patients per ICD-9 chapter histogram

Patients distribution across ICD-9 chapters (multiple counts possible).

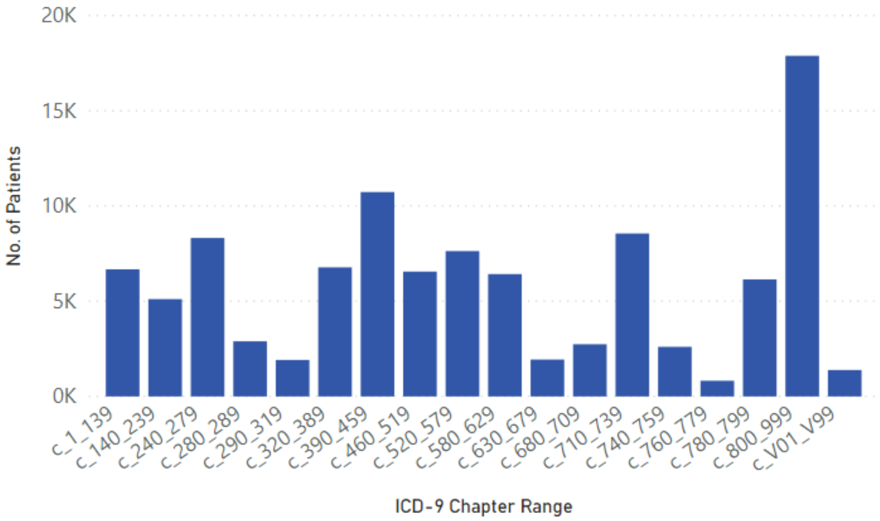


Figure 6 (continued): ICD-9 chapter legend

Each chapter corresponds to a range of ICD-9 codes.

ICD-9 Chapter Range	**ICD-9 Chapter Description**
c_1_139	Infectious And Parasitic Diseases
c_140_239	Neoplasms
c_240_279	Endocrine, Nutritional And Metabolic Diseases, And Immunity Disorders
c_280_289	Diseases Of The Blood And Blood-Forming Organs
c_290_319	Mental Disorders
c_320_389	Diseases Of The Nervous System And Sense Organs
c_390_459	Diseases Of The Circulatory System
c_460_519	Diseases Of The Respiratory System
c_520_579	Diseases Of The Digestive System
c_580_629	Diseases Of The Genitourinary System
c_630_679	Complications Of Pregnancy, Childbirth, And The Puerperium
c_680_709	Diseases Of The Skin And Subcutaneous Tissue
c_710_739	Diseases Of The Musculoskeletal System And Connective Tissue
c_740_759	Congenital Anomalies
c_760_779	Certain Conditions Originating In The Perinatal Period
c_780_799	Symptoms, Signs, And Ill-Defined Conditions
c_800_999	Injury And Poisoning
c_V01_V99	Supplementary Classification Of Factors Influencing Health Status And Contact With Health Services

Figure 7: Injury and poisoning sub-chapter histogram

Patients distribution across injury and poisoning sub-chapters (multiple counts possible).

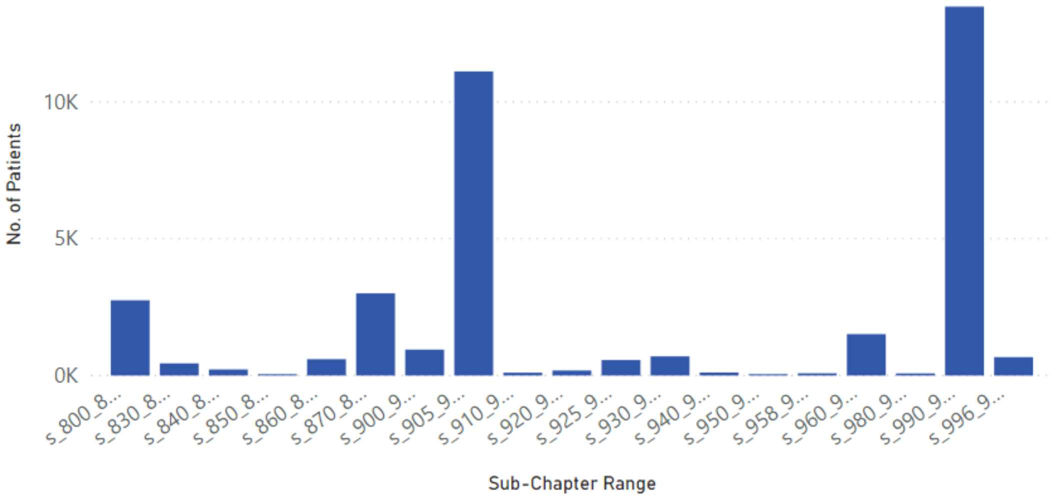


Figure 7 (continued): Injury and poisoning sub-chapter legend

Each sub-chapter corresponds to a range of ICD-9 codes.

ICD-9 Sub-chapter Range	ICD-9 Sub-chapter Description
s_800_829	Fractures
s_830_839	Dislocation
s_840_848	Sprains And Strains Of Joints And Adjacent Muscles
s_850_854	Intracranial Injury, Excluding Those With Skull Fracture
s_860_869	Internal Injury Of Thorax, Abdomen, And Pelvis
s_870_897	Open Wounds
s_900_904	Injury To Blood Vessels
s_905_909	Late Effects Of Injuries, Poisonings, Toxic Effects, And Other External Causes
s_910_919	Superficial Injury
s_920_924	Contusion With Intact Skin Surface
s_925_929	Crushing Injury
s_930_939	Effects Of Foreign Body Entering Through Orifice
s_940_949	Burns
s_950_957	Injury To Nerves And Spinal Cord
s_958_959	Certain Traumatic Complications And Unspecified Injuries
s_960_979	Poisoning By Drugs, Medicinal And Biological Substances
s_980_989	Toxic Effects Of Substances Chiefly Nonmedicinal As To Source

s_990_995	Other And Unspecified Effects Of External Causes
s_996_999	Complications Of Surgical And Medical Care, Not Elsewhere Classified

Figure 8: Diseases of the circulatory system sub-chapter histogram

Patients distribution across diseases of circulatory system sub-chapters (multiple counts possible).

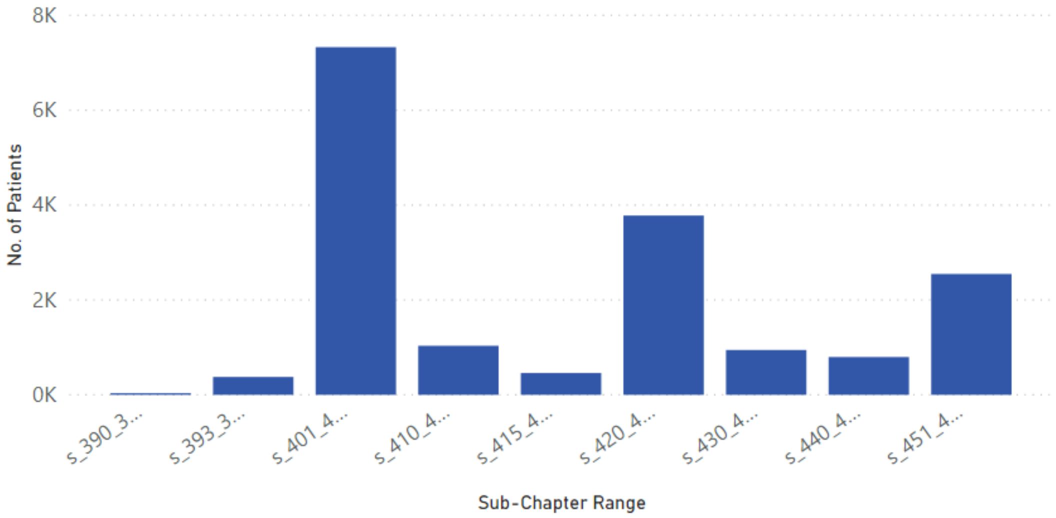


Figure 8 (continued): Diseases of the circulatory system sub-chapter legend

Each sub-chapter corresponds to a range of ICD-9 codes.

ICD-9 Sub-chapter Range	ICD-9 Sub-chapter Description
s_390_392	Acute Rheumatic Fever
s_393_398	Chronic Rheumatic Heart Disease
s_401_405	Hypertensive Disease
s_410_414	Ischemic Heart Disease
s_415_417	Diseases Of Pulmonary Circulation
s_420_429	Other Forms Of Heart Disease
s_430_438	Cerebrovascular Disease
s_440_449	Diseases Of Arteries, Arterioles, And Capillaries
s_451_459	Diseases Of Veins And Lymphatics, And Other Diseases Of Circulatory System

Figure 9: ICD-9 hierarchy embedded in 2-d space

Built on synthetic data, the ICD-9 hierarchy is reconstructed in the Poincaré disk model and then transferred to Euclidean space (Beaulieu-Jones et al., 2019).

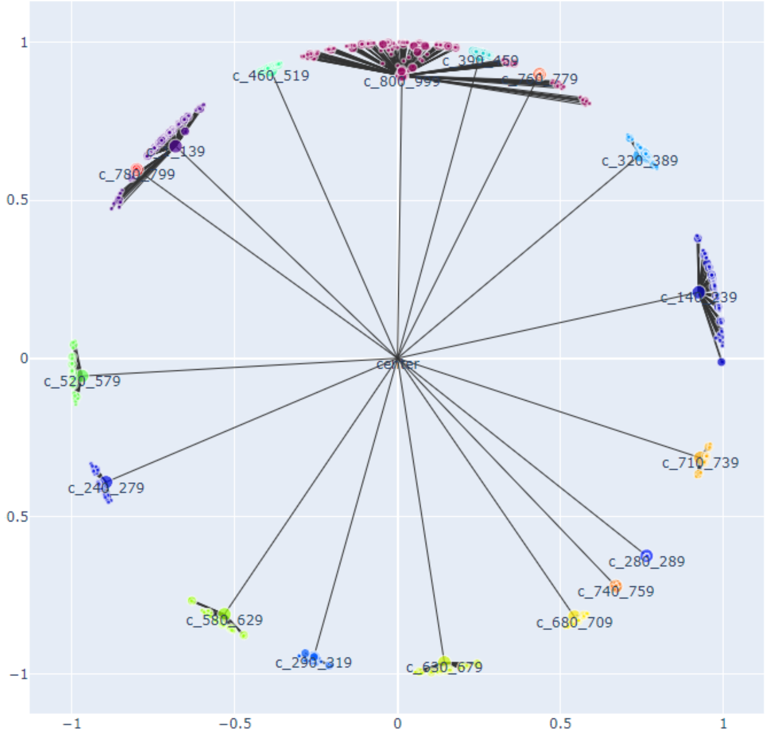
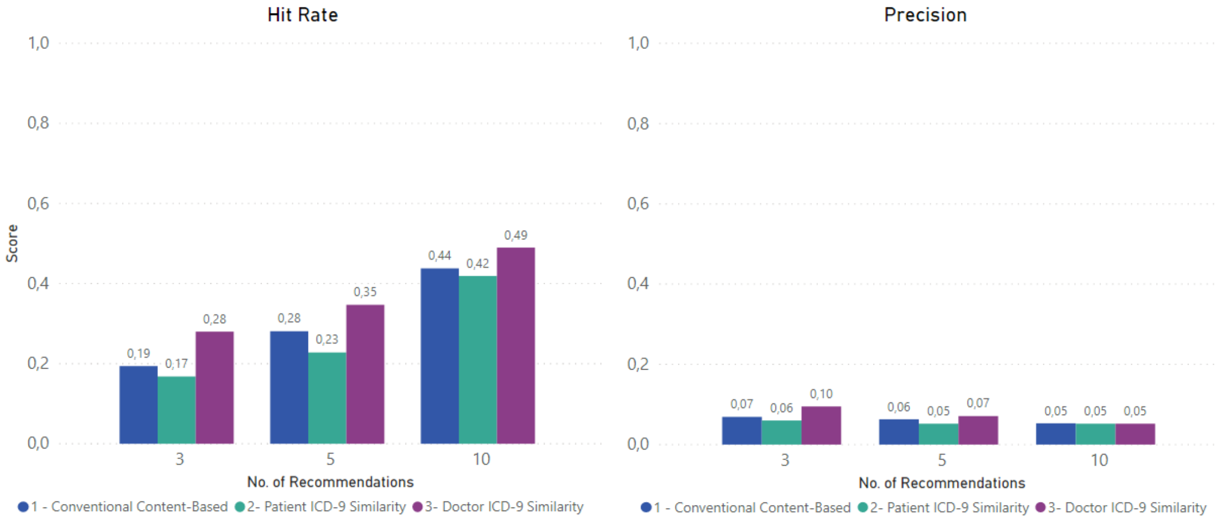


Figure 10: Model evaluation

Hit rate and precision scores per proposed model.



B. Tables

Table 1: Summary statistics of 201912 release of US NLM ICD-9 to SNOMED CT map

Table indicating how many ICD-9 codes have a unique, an ambiguous or no mapping to the SNOMED CT (US NLM, 2019).

Map Type	**ICD-9 Codes**	**% of usage of ICD-9 codes**
1-1 Maps	7,826 (66%)	70%
1-M Maps	3,350 (28%)	25%
No Map	672 (6%)	5%
Total	**11,848 (100%)**	**100%**

C. Equations

Equation (1): Hyperbolic distance

$$d_H(x,y) = \text{acosh}\left(1 + 2\frac{||x-y||^2}{(1-||x||^2)(1-||y||^2)}\right)$$

Equation (2): Distance between two children in graph in continuous space

$$d(x,y) = d(x,O) + d(O,y)$$

Equation (3): Distance ratio of original input graph (normalisation of equation (2))

$$\frac{d(x,y)}{d(x,O) + d(O,y)} = 1$$

Equation (4): Cosine similarity

$$\cos(\theta) = \frac{A * B}{||A|| * ||B||}$$

Equation (5): Pearson correlation

$$\rho(A,B) = \frac{cov(A,B)}{\sigma A \sigma B}$$

Equation (6): Hyperbolic Average (Einstein midpoint)

$$HypAve(x_1, \dots, x_N) = \frac{\sum_{i=1}^{N} \gamma_i x_i}{\sum_{i=1}^{N} \gamma_i}$$

$$with\, \gamma_i = \frac{1}{\sqrt{1 - c||x_i||^2}}\ and\ c = 1$$

Equation (7): Feature vectors of patient *i* and doctor *j*:

$$f_i = HypAve(V_i)$$

$$f_j = HypAve(V_j)$$

Equation (8): Hyperbolic similarity between doctors *j* and *k*:

$$s_{j,k} = s_H(f_j, f_k)$$

Equation (9): Predicted affinity of patient *i* towards a doctor *j*:

$$p_{i,j} = \frac{\sum_{k=1}^{K} y_{i,k} * s_{j,k}}{\sum_{k=1}^{K} s_{j,k}}$$