Populating Physician Biographical Pages Based on EMR Data

Feichen Shen, Ph.D.[1], Sunghwan Sohn, Ph.D.[1], Majid Rastegar-Mojarad, M.S.[1], Sijia Liu, M.S.[1], Joshua J. Pankratz M.S.[2], Michael A. Hatton[3], Nancy Sowada[3], Om K. Shrestha, M.B.A.[4], Shawna L. Shurson M.B.A.[4], Hongfang Liu, Ph.D.[1]
[1]Department of Health Sciences Research [2]Division of Information Management and Analytics [3]Global Business Solutions [4]Enterprise Analytics
Mayo Clinic, Rochester, MN, USA

Abstract

The physicians' biographical pages are essential in providing information about physicians' specialties. However, physicians may not have biographical pages or the current pages are not comprehensive. We hypothesize that physicians' specialty information can be mined from Electronic Medical Records (EMRs) of their patients. We proposed an automated physician specialty populating (PSP) system that analyzes physician-ascertained diagnoses in EMRs, aggregates them to an appropriate granularity based on the current biographical pages, and populates the biographical pages accordingly. In this study, we applied the system using EMR data from Mayo Clinic and evaluated the system using the current biographical pages regarding various ranking strategies. Preliminary results demonstrated that using EMR data is a scalable and systematic way to populate physicians' biographical pages.

Introduction

With the wide usage of information technology in healthcare, physicians' biographical pages have become the most commonly accessed institutionally oriented web pages. An internal study conducted by the Mayo Clinic User Experience (UX) team indicated that biographical information facilitates decision-making of key audiences, including current and prospective patients, referring physicians, research collaborators, students, job seekers, and grant-funding agencies. The study also identified the specific types of biographical information needed to support decision-making by different audiences.

Enriching physicians' biographical pages is critical in healthcare delivery as it can align "people, processes, data and technologies to optimize information, collaboration, expertise, and experience" according to Healthcare Information & Management Systems Society (HIMSS)[1]. Especially in the era of Electronic Medical Records (EMRs) and big data, data-driven decision making has been widely adopted in the healthcare domain. Lobach et al., have conducted experiments to show strong evidences of using both clinical decision support systems (CDSSs) and knowledge management systems (KMSs) to facilitate the process of healthcare management[2]. Facing a large amount of under-utilized data stored in hospital systems, Chawla and Davis proposed a patient individualized framework to improve personalized healthcare[3]. Similarly, Dixon et al., developed a cloud based distributed framework for knowledge management and clinical decision support[4]. Furthermore, by utilizing the combination of local mining and global learning approach, Nie et al., is able to bridge the vocabulary gap between health consumers and healthcare knowledge[5]. In addition, some other researches incorporate with semantic web technology to support decision making, play as an active role in managing healthcare knowledge[6-8].

There are some expert recommendation systems to impetus healthcare delivery. For example, Jiang and Xu proposed a doctor recommendation model which includes a quality analysis component[9]. Fang and Zhai targeted on the problem of expert finding by adopting probabilistic models with the focus on building ranking oriented match making between candidate profiles and topic groups[10]. Balog et al., applied a language modeling framework to achieve expert finding[11]. These studies focus on intelligent interfaces for consumers at the point of care seeking not about how to populate the content of biographical pages.

Hence, there is an urgent need to have physicians' biographical pages populated with useful content at the point of needs for diverse information seekers. However, biographical pages are usually filled up manually and displayed in various types of design, standards, content, and functionality. Furthermore, take Mayo Clinic as an example, although there are 88.7% staff physicians and scientists having biographical pages across three campuses (Arizona, Florida and Minnesota), only 72.7% of them have their interests field populated with their interests and specialties. Many healthcare organizations do not maintain a comprehensive list of specialties of their physicians. One major reason is the lack of time for physicians to create and update their biographical pages. Additionally, biographical

pages are not well-integrated with other contents from other resources, indicating a missing opportunity in facilitating decision making.

In this paper, we proposed an automated physician specialty populating (PSP) system. Our hypothesis is that the diagnosis information in EMRs represents the specialties of the corresponding physicians and it can be utilized to populate physicians' biographical pages. The paper is organized as follows: we first introduce methodology and workflow of our system. Then we describe experiments and present the results. Finally, current issues and possible ways to make improvement as future directions are discussed.

Methods

Our PSP system is composed of multiple components as shown in Figure 1. The system starts with the component of EMRs parsing and normalization, which performs as a preprocessing step on unstructured EMRs so that only keep diagnosis sections for physicians. The significant diagnoses detection component takes outputs from the previous step to filter out non-significant diagnoses and keep those with high frequencies for each specific physician. The semantic granularity checking component reviews high-ranked diagnoses for each physician and maps them to SNOMED-CT. Meanwhile, this component also takes the existing biographical pages as gold standard and annotates them with SNOMED-CT in order to check whether EMR diagnoses could be mapped to gold standard terms from a semantic hierarchical perspective to finalize the granularity. In the rest of this section, we provide detailed strategies, methods and algorithms for each component.
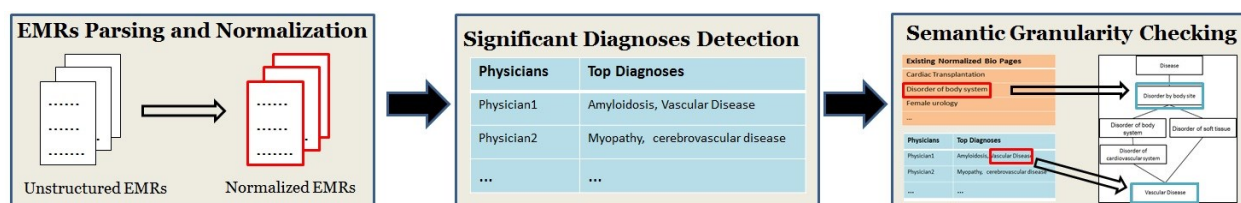

Figure 1. Workflow of the proposed physician specialty populating (PSP) system

EMR Parsing and Normalization

This component retrieves clinical notes and only uses physician-ascertained diagnoses under the diagnosis section. Along with diagnoses, we also prepare specific patient ID, practice setting, provider ID and provider type. For the purpose of normalization, we tokenize sentences of diagnoses and process them through MetaMap[12] to get relevant UMLS concepts, preferred names and concept unique identifiers (CUIs)[13]. To better improve the accuracy of specialty populating, we put a restriction on selecting CUIs by checking their semantic types. As shown in Table 1, we only focus on CUIs with semantic type *patf*, *dsyn*, *mobd*, *neop*, *comd*, *emod*, *hlca*, *lbpr*, *diap*, and *topp*, which are specific to diseases and procedures.

In addition, to represent diagnoses in ontology, we conduct a further annotation about converting CUIs to SNOMED-CT codes[14, 15] by using UMLS Metathesaurus file in Rich Release Format (RRF)[13, 16]. Particularly, to address the mismatching issue between CUIs and SNOMED-CT codes, we not only find mapping codes directly in Metathesaurus RRF file, but also check the concepts/preferred names and make an algorithm to optimize the decision-making in a heuristic way. Basically, we first conduct a mapping between CUIs and SNOMED-CT codes. If there is only one matching, we directly use the outcome. But if there are more than one matched codes, it means there exists a possibility of mismatch. To make this step accurately done, we check Metathesaurus RRF file with concept/preferred names to retrieve SNOMED-CT codes directly from there. As a result, outputs generated from this preprocessing component are a list of physicians with several SNOMED-CT codes along with patients' information, practice settings and providers' details.

Table 1. Qualified semantic types

UMLS Semantic Type	Type Unique Identify (TUI)	Abbreviation
Pathologic Function	T046	patf
Disease or Syndrome	T047	dsyn
Mental or Behavioral Dysfunction	T048	mobd

Neoplastic Process	T191	neop
Cell or Molecular Dysfunction	T049	comd
Experimental Model of Disease	T050	emod
Health Care Activity	T058	hlca
Laboratory Procedure	T059	lbpr
Diagnostic Procedure	T060	diap
Therapeutic or Preventive Procedure	T061	topp

Significant Diagnoses Detection

This component is responsible for selecting top diagnoses for each physician. To achieve this goal, we respectively apply three algorithms on SNOMED-CT codes for all physicians: 1) term frequency (TF); 2) term frequency-inverse document frequency (TF-IDF) [17]; 3) a hybrid algorithm using TF and TF-IDF. TF based algorithm simply picks the top K frequent SNOMED-CT codes and considers them as the most significant representatives. The TF-IDF aims to weight how important a term is to a document within a number of documents. In this study, we treat each physician as a document, and consider all SNOMED-CT codes as terms. Therefore, by using TF-IDF, we are able to find more important SNOMED-CT codes related to each physician and consider them as significant ones. The process of TF-IDF can be summarized as:

$$tf(c,p) = \frac{n_{c,p}}{\sum_i n_{i,p}} \qquad (Eq\ 1)$$

$$idf(c,P) = log\frac{P}{|\{p \in P : c \in p\}|} \qquad (Eq\ 2)$$

$$tfidf(c,p,P) = tf(c,p) \cdot idf(c,P) \quad (Eq\ 3)$$

where $tf(c,p)$ is weighted term frequency for each SNOMED-CT code c in each physician p, $n_{c,p}$ stands for the number of occurrences of the considered code c for physician p, $idf(c,P)$ represents inverse document frequency for code c in the whole physician collection P, and $tfidf(c,p,P)$ is calculated by the product of $tf(c,p)$ and $idf(c,P)$. Based on the weight measured by TF-IDF, we choose the top K SNOMED-CT codes for each physician and rank them in a descending order as the output of this component.

We also combine TF and TF-IDF approaches and select the top A terms based on TF and the top B terms based on TF-IDF. In our implementation, we choose A and B equally.

Semantic Granularity Checking

Diagnoses in clinical notes are always described in a specific way. However, patients may have limited medical knowledge where specific terms may not have the right granularity for patients. In this step, we identify the appropriate granularity based on the current biographical pages.

Specifically, as the current biographical pages are manually written by physicians about their high level specialties and general interests, it can be served as a gold standard to determine the appropriate granularity and evaluate the ranking algorithms. Here, we map unstructured terms contained in biographical pages to SNOMED-CT codes with the annotation approach described in the EMR Parsing and Normalization component.

Then we use information content[18] to check hierarchical granularity of each concept in ontology. Information content in ontology illustrates how informative one node is based on its annotation frequencies and how many descendants it holds. Specifically for SNOMED-CT, we only consider two nodes as ancestors and descendants relationship if they are connected with "is-a" relationship. The measurement of SNOMED-CT annotation starts with a probability measure of each SNOMED-CT term t. Let S_t be the collection of SNOMED-CT terms that are either t or its

descendants and u be the element in S_t. Let O *(t, c)* be the occurrence of t annotations given a collection c. The information probability of t in c, denoted as p *(t, c)*, is defined as Equation 4 shown[19].

$$p\ (t,c) = \frac{O(t,c)}{\sum_{u \in S_t} O(u,c)} \quad (Eq\ 4)$$

Figure 2 gives an example to illustrate how information content is calculated. In this graph, as bottom level nodes, concept 7200100 occurs 126 times, concept 9813009 occurs 4 times and concept 37119500 occurs 33 times. Based on "is-a" link, frequencies appear on bottom level can be propagated to their direct parent nodes, which makes frequency of concept 281244004 become 126, and frequency of concept 281243005 become 33 + 4 = 37. Similarly, as the top level in this example, concept 281242000 holds a total frequency as 126 + 37 = 163. It is obvious that the frequency and probability of each concept increases as we move up the graph to the root, and the probability of root concept will be 1. Therefore, we can quantify the hierarchical relationship based on information probability, and granularity is able to be determined according to the probability threshold we set.
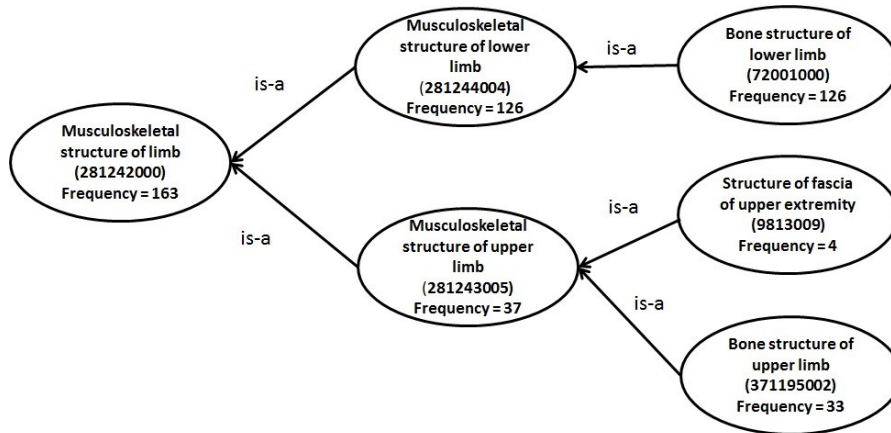

Figure 2. An example of information content measurement in SNOMED-CT

Considering both SNOMED-CT hierarchy and information content, we propose an algorithm to semantically check granularity as shown in Algorithm 1. For each diagnosis SNOMED code, we first check if gold standard has the exactly same code. If not, we check whether gold standard has ancestors of such concept within the depth of θ. Otherwise, we compare information probability of the specific node with a granularity threshold φ. If information content meets the criteria, even though there is no match in gold standard, we still consider such node as a concept with high granularity.

Algorithm 1. Semantic Granularity Checking

Input: Map A (physician, List of SNOMED codes), List B of SNOMED codes in physician's biography, level threshold θ, granularity threshold φ

Output: Map C (physician, List of Promoted SNOMED codes)

1. **FOR** each physician p in A

2.   **FOR** each SNOMED code c

3.   **IF** B has the exact same code as c

4.   add c to List T

5.   **ELSE IF** B contains c's ancestor with k levels ($1<k<\theta$)

6.   add c to List T

7.   **ELSE IF** information probability ip$> \varphi$

8.   add c to List T

---

9. add (p, T) to C

10. **RETURN** C

Experimental Results

This proposed system was implemented with Eclipse Standard/SDK version Luna 4.4.0[20]. The interfaces to access MetaMap, UMLS and SNOMED-CT were coded in Java programming language. To facilitate the processing speed on large amount of clinical notes, we also deployed the PSP system on the Open Grid Scheduler (OGS)[21] framework running on 64 bit Linux CentOS 6.8 servers hosted by Mayo Clinic, which is a scheduler to run distributed tasks on clusters.

The clinical notes we processed were collected during the year of 2010 to 2015 provided by Mayo Clinic Minnesota campus. Table 2 gives detailed statistics of the data. To accurately find specialties for each physician, we processed a subset of clinical notes to keep problems of all patients which first diagnosed by each physician, which ends up with the same amount of patients but lower number of providers and clinical notes. Among these clinical notes, we parsed all problem lists contained in diagnosis section and stored them along with patients' and providers' information. One point needs to be noticed is that, although there are 8,078 providers showed up on EMRs from Mayo Clinic Minnesota campus, since Mayo Clinic only set up biographical pages for doctors who serve as primary or secondary appointment, therefore, not all of those providers have clinical biographies. As shown in Table 3, from Mayo Clinic Arizona, Florida and Minnesota campuses, there are 2,967 physicians currently have clinical biographies, and 2,431 out of 2,967 (81.9%) physicians have their interests fields populated on clinical biographies, which we considered as gold standard. In addition, among these 2,431 physicians, 718 physicians can be found from EMRs of Minnesota campus. As a comparison, we also populated physicians' specialties by analyzing patients' claim data, which describes patients' billing information submitted by physicians and hospitals. We retrieved 2010-2015 billing data from all three campuses of Mayo Clinic with 3,002 physicians and found that only 658 out of 718 physicians from Minnesota campus have both EMRs and relevant claim data. As a case study, the experiments were conducted based on those 658 physicians.

For those 658 physicians, we extracted the top 100 significant SNOMED-CT codes based on the three algorithms. We also used EMRs, claim data, and the combination of EMRs and claim data to evaluate the performance. Then we applied the granularity checking algorithm to map each SNOMED code to an appropriate level. Specifically, we set parameter θ as 6, and threshold φ as the average of the information probability of annotated SNOMED-CT codes. In general, nine different experimental groups are formed as: 1) TF-IDF with EMRs; 2) TF-IDF with claim data; 3) TF-IDF with EMRs and claim data; 4) TF with EMRs; 5) TF with claim data; 6) TF with EMRs and claim data; 7) Hybrid with EMRs; 8) Hybrid with claim data; 9) Hybrid with EMRs and claim data.

Table 2. Statistics of Mayo Clinic 2010-2015 EMRs

Cohorts	Number of Patients	Number of Providers	Number of Clinical Notes
EMRs	789,966	8,249	23,979,937
Subset of EMRs	789,966	8,078	16,094,797

Table 3. Statistics of current physicians' clinical biographies from Mayo Clinic three campuses

Data	Number of Physicians with Bio	Number of Physicians with Interests in Bio	Number of Physicians with Mayo Clinic Minnesota EMR	Number of Physicians with Mayo Clinic Minnesota EMR and Claim Data
Current Bio Pages	2,967	2,431	718	658

The total number of unique SNOMED-CT codes is 999, 1,038, 1,244, 355, 330, 319, 1,564, 1,216, and 1,245 for experimental group 1 to group 9, respectively. In addition, we grouped physicians by their practice settings and

counted the number of SNOMED-CT codes for each category with different experimental groups. As shown in Figure 3, one physician may have one or more practice settings, and we listed the top 10 practice settings with the descending order of population of physicians in x axis with the number of physicians in parentheses. *Cardiology* had more physicians, and *Nephrology* had more SNOMED-CT codes than other settings. The combination of EMRs and claim data always produced the most number of SNOMED-CT codes while running on claim data always led to the least number of codes.

In Figure 4, we elaborated the precision (rounded to two decimal places) of SNOMED-CT codes that covered by the gold standard for these 10 practice settings for nine different experimental groups. The claim data yields low precision. However, the selection of different approaches to acquire high precision varies across different practice settings. Additionally, the current biographic pages do not seem to represent the practice of the physicians as codes mined from EMR are mostly missed from their current pages. For example, a physician has *Cardiac transplantation* and *Mechanical circulatory support* listed in their biographic page. However, we found more practical procedures including *Aortic valve stenosis* and *Mitral stenosis* in his clinical notes. This may reflect the gap between what physicians think their specialties/research and what they are actually doing in routine practice. The PSP system is able to quantify such gap for physicians and enrich their biographical pages with their clinical specialties.
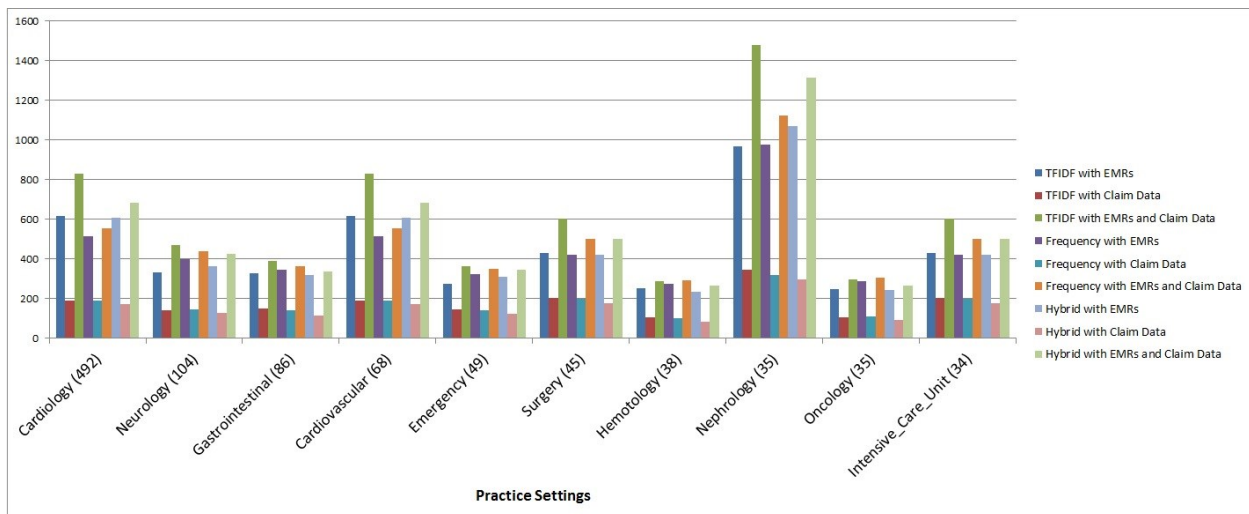

Figure 3. Count of SNOMED-CT codes for the top 10 practice settings with 9 experimental configurations

We conducted a comparison among 9 experimental groups in terms of their performance. As shown in Table 4, overall precision/recall/F-measure results for 658 physicians were given. We found that the use of TF-IDF algorithm with EMRs produced the highest precision as 0.49, the use of TF-IDF algorithm with EMRs and claim data contributed to the highest recall as 0.67, and the use of hybrid algorithm with EMRs yielded highest F-measure as 0.51. It showed that EMRs is able to provide more significant specialties than claim data with either TF-IDF or hybrid algorithm, while using frequency algorithm was not good enough to generate as informative contents as the others. However, for claim data, no matter what kind of algorithms we applied, it always performed with low precision/recall/F-measure, it indicated that claim data did not cover a fine granularity of specialties and cannot reflect physicians' practices very well. Moreover, the combination of EMRs and claim data did not have a better performance than using EMRs. It is because EMRs contain more useful information than what claim data holds, and when we mixed them together, the ranking of terms was changed, and less informative term may have a higher frequency, which makes important terms become invisible.

Although we got the highest F-measure by using hybrid algorithm with EMRs, 0.51 is not satisfactory score. The reason is that gold standard we referred to did not have a comprehensive description of practices while most diagnose sections EMRs provide are about clinical practices and procedures. In other words, F-measure score with 0.51 actually quantified such gap between clinical practices and real specialties. Ideally, with more practices involved in biographical pages, higher F-measure score will be achieved.

Table 4. Overall precision/recall/F-measure for 658 physicians with nine groups (highest value in bold)

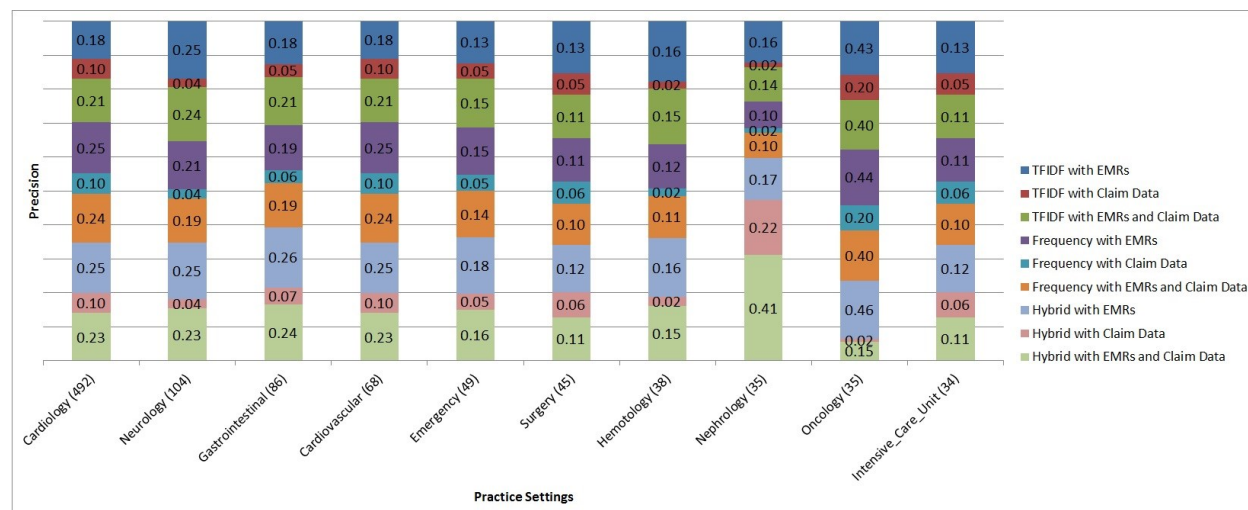| Methods	EMRs	Claim Data	EMRs + Claim Data
TF-IDF	**0.49**/0.52/0.5	0.21/0.08/0.12	0.4/**0.67**/0.5
TF	0.45/0.5/0.47	0.21/0.07/0.11	0.42/0.54/0.47
Hybrid	0.45/0.6/**0.51**	0.22/0.07/0.11	0.42/0.63/0.5

Figure 4. Precision for the top 10 practice settings with 9 experimental configurations

Discussion

Our proposed physician specialty populating (PSP) system is able to automatically populate potential specialty candidates for physicians from EMRs. The motivation of making this system is based on observations that most physicians fill their biographical pages manually with no standardization, which makes it difficult to deliver healthcare knowledge to consumers and providers. In addition, physicians' biographical pages may not be up to date. The proposed PSP system has the potential to enrich the current biographical pages and keep it up to date by leveraging EMR.

Normalizations for both EMRs and existing biographical pages are initial processes of the proposed system. Note that here, we took existing biographical pages as gold standard, the coverage of 60.4% may make the granularity checking suboptimal.

There is a debate of whether using TF or TF-IDF to find significant SNOMED-CT codes for physicians from EMRs. Usually, frequency of SNOMED-CT codes tells us which specialties are commonly applied by each physician and TF-IDF can indicate which specialties are especially unique for each physician. From patients' point of view, frequency is an essential key since it tells them which specialties physicians are frequently performing. TF-IDF also provides specialties information with not only relative high frequency but also uniqueness for each physician. Such ranking scheme will rank more specific specialties to a specific physician higher. However, using TF-IDF might miss some common practices that with high frequencies but most physicians can do. In this study, we gave the same weight to each approach aiming to get benefit from both approaches. Although the hybrid approach yielded the best F-measure, there is still some room for improvement. In the future, we will use a heuristic approach to find the optimal balance point in combining the two approaches. Meanwhile, we will also send outputs to physicians at Mayo Clinic to get their feedback to adjust the weight. By doing so, we aim to make better balance between consumers' needs and physicians' requirements.

One challenge we faced is how to decide the right granularity. We always wanted to give coarse granularities with more general terms for the convenience of search. However, terms with fine granularities may also be useful to provide specific division of specialties. Therefore, in this study, we picked the granularity based on the average of the information content as the threshold to balance the granularity. Note that, the PSP system tends to populate what

physicians do in the practice. From the experiments, we found that claim data cannot reflect physicians' specialties very well which may be due to the institutional and financial constraints. Hence, it provides evidences to the current situation of health systems that there exists a divergence between structured claim data and unstructured EMRs. How to use both datasets to understand the status and activities of healthcare is another interesting topic to investigate on.

Conclusion

We have investigated on the use of EMR data to populate physicians' biographical pages. The proposed PSP system provides a scalable way to automatically populate physicians' biographical pages. We plan to collaborate with physicians and refine system settings. In addition to populating physicians' clinical specialties, the same approach will be utilized to discover their research expertise by analyzing keywords appeared in their research articles and grant proposals.

Acknowledgement

References

1. Guptill J. Knowledge management in health care. J Health care finance. 2005 Apr;31(3):10-14.
2. Lobach D, Sanders GD, Bright TJ, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux R, Samsa G, Hasselblad V, Williams JW. Enabling health care decisionmaking through clinical decision support and knowledge management. Evidence report/technology assessment. 2012 April.
3. Chawla NV, Davis DA. Bringing big data to personalized healthcare: a patient-centered framework. Journal of general internal medicine. 2013 Sep 1;28(3):660-665.
4. Dixon BE, Simonaitis L, Goldberg HS, Paterno MD, Schaeffer M, Hongsermeier T, Wright A, Middleton B. A pilot study of distributed knowledge management and clinical decision support in the cloud. Artificial intelligence in medicine. 2013 Sep 30;59(1):45-53.
5. Nie L, Zhao YL, Akbari M, Shen J, Chua TS. Bridging the vocabulary gap between health seekers and healthcare knowledge. IEEE Transactions on Knowledge and Data Engineering. 2015 Feb 1;27(2):396-409.
6. Hussain S, Abidi SR, Abidi SS. Semantic web framework for knowledge-centric clinical decision support systems. In Conference on Artificial Intelligence in Medicine in Europe. Springer Berlin Heidelberg. 2007 Jul 7;451-455.
7. Abidi SR, Hussain S, Shepherd M, Abidi SS. Ontology-based modeling of clinical practice guidelines: a clinical decision support system for breast cancer follow-up interventions at primary care settings. In Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems. IOS Press. 2007;845.
8. Dieng-Kuntz R, Minier D, Růžička M, Corby F, Corby O, Alamarguy L. Building and using a medical ontology for knowledge management and cooperative work in a health care network. Computers in Biology and Medicine. 2006 Aug 31;36(7):871-892.
9. Jiang H, Xu W. How to find your appropriate doctor: An integrated recommendation framework in big data context. In Computational Intelligence in Healthcare and e-health (CICARE), 2014 IEEE Symposium on 2014 Dec 9;154-158.
10. Fang H, Zhai C. Probabilistic models for expert finding. The 29th European Conference on Information Retrieval. 2007 Apr 2; 418-430.
11. Balog K, Azzopardi L, de Rijke M. A language modeling framework for expert finding. Information Processing & Management. 2009 Jan 31;45(1):1-9.
12. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. In Proceedings of the AMIA Symposium. American Medical Informatics Association. 2001;17.
13. Schuyler PL, Hole WT, Tuttle MS, Sherertz DD. The UMLS Metathesaurus: representing different views of biomedical concepts. Bulletin of the Medical Library Association. 1993 Apr;81(2):217.
14. Cote RA, College of American Pathologists. Systematized nomenclature of medicine. College of American Pathologists; 1977.
15. Donnelly K. SNOMED-CT: The advanced terminology and coding system for eHealth. Studies in health technology and informatics. 2006 Jan;121:279.

16. Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic acids research. 2004 Jan 1;32(suppl 1):267-270.

17. Luhn HP. A statistical approach to mechanized encoding and searching of literary information. IBM Journal of research and development. 1957 Oct;1(4):309-317.

18. Lord PW, Stevens RD, Brass A, Goble CA. Investigating semantic similarity measures across the Gene Ontology: the relationship between sequence and annotation. Bioinformatics. 2003 Jul 1;19(10):1275-1283.

19. Liu H, Hu ZZ, Wu CH. DynGO: a tool for visualizing and mining of Gene Ontology and its associations. BMC bioinformatics. 2005 Aug 9;6(1):1.

20. Eclipse Luna. Available: https://eclipse.org/luna/. Accessed 7 September 2016.

21. Oracle and Sun Microsystems. Available: https://www.oracle.com/sun/index.html. Accessed 7 September 2016