International Journal of Healthcare Sciences ISSN 2348-5728 (Online)
Vol. 3, Issue 1, pp: (228-232), Month: April 2015 - September 2015, Available at: www.researchpublish.com

Ontology Based Medical Terminology System

Vinoliyavinamuthu A.

Master of computer applications, V.S.B.Engineering College, Karur-639111, India

Abstract: **In the last two decades we have witnessed considerable efforts directed towards making electronic healthcare records comparable and interoperable through advances in record architectures and (bio) medical terminologies and coding systems. Deep semantic issues in general and ontology in particular, have received some interest from the research communities. However, with the exception of work on so-called 'controlled vocabularies', ontology has thus far played little role in work on standardization. The prime focus has been rather the rapid population of terminologies at the level of fine detail. In this paper, we argue that more efforts are needed on the side of both research and standardization to ensure that the coding systems used in electronic healthcare records enjoy a semantics that is coherent with the semantics of the record. We propose realist ontology as a method to bring about this coherence by means of a robust system of top-level ontological categories.**

Keywords: **ontology, semantics**

1. INTRODUCTION

Information technologies are transforming the ways healthcare services are delivered, from patients' passively embracing their doctors' orders to patients' actively seeking online information that concerns their health. For professionals, they are able to increase their reputations among their colleagues and patients, strengthen their practical knowledge from interactions with other renowned doctors, as well as possibly attract more new patients. For patients, these systems provide nearly instant and trusted answers especially for complex and sophisticated problems. Over times, a tremendous number of medical records have been accumulated in their repositories, and in most circumstances, users may directly locate good answers by searching from these record archives, rather than waiting for the experts' responses or browsing through a list of potentially relevant documents from the Web. In many cases, the community generated content, however, may not be directly usable due to the vocabulary gap. Users with diverse backgrounds do not necessarily share the same vocabulary. Take Health Tap as an example, which is a question answering site for participants to ask and answer health-related questions. The questions are written by patients in narrative language.

The same question may be described in substantially different ways by two individual health seekers. On the other side, the answers provided by the well-trained experts may contain acronyms with multiple possible meanings, and non-standardized terms. Recently, some sites have encouraged experts to annotate the medical records with medical concepts. However, the tags used often vary wildly and medical concepts may not be medical terminologies. For example, "heart attack" and "myocardial disorder" are employed by different experts to refer to the same medical diagnosis. It was shown that the inconsistency of community generated health data greatly hindered data exchange, management and integrity.

Even worse, it was reported that users had encountered big challenges in reusing the archived content due to the incompatibility between their search terms and those accumulated medical records. Therefore, automatically coding the medical records with standardized terminologies is highly desired. It leads to a consistent interoperable way of indexing, storing and aggregating across specialties and sites.

In addition, it facilitates the medical record retrieval via bridging the vocabulary gap between queries and archives. It is worth mentioning that there already exist several efforts dedicated to research on automatically mapping medical records to terminologies. Most of these efforts, however, focused on hospital generated health data or health provider released sources by utilizing either isolated or loosely coupled rule-based and machine learning approaches. Compared to these kinds of data, the emerging community generated health data is more colloquial, in terms of inconsistency, complexity and ambiguity, which pose challenges for data access and analytics.

Further, most of the previous work simply utilizes the external medical dictionary to code the medical records rather than considering the corpus-aware terminologies. Their reliance on the independent external knowledge may bring in inappropriate terminologies. Constructing a corpus- aware terminology vocabulary to prune the irrelevant terminologies of specific dataset and narrow down the candidates is the tough issue we are facing. In addition, the varieties of heterogeneous cues were often not adequately exploited simultaneously. Therefore, a robust integrated framework to draw the strengths from various resources and models is still expected.

2. LITERATURE SURVEY

A Joint Local-Global Approach for Medical Terminology Assignment: The rise of digital technologies has transformed the patient-doctor relationships. Nowadays, when patients struggle with their health concerns, the majority usually explore the Internet to research the problem before and after they see their doctors. For example, 70% of Canadians turned to Internet to look up health-related information and 72% of American Internet users searched for health solutions in. These metrics have reflected the scope and scale of the online health seekers. To better serve the needs of health seekers, communitybased health services have emerged as effective platforms for health knowledge dissemination and exchange, such as HealthTap1 , HaoDF2 and WenZher. They not only permit health seekers to freely post health-oriented questions, but also encourage doctors to provide trustworthy answers. Over time, a tremendous number of QA pairs has been accumulated in their repositories, and in most circumstances, health seekers may directly locate good answers by searching from these archives, rather than waiting for the experts' responses or painfully browsing through a list of documents from the general search engines. In community-based health services, vocabulary gap between health seekers and community generated knowledge has hindered data access. To bridge this gap, this paper presents a scheme to label question answer(QA) pairs by jointly utilizing local mining and global learning approaches. Local mining attempts to label individual QA pair by independently extracting medical concepts from the QA pair itself and mapping them to authenticated terminologies. However, it may suffer from information loss and lower precision, which are caused by the absence of key medical concepts and presence of irrelevant medical concepts. Global learning, on the other hand, works towards enhancing the local mining via collaboratively discovering missing key terminologies and keeping off the irrelevant terminologies by analyzing the social neighbors. Practically, this unsupervised scheme holds potential to large-scale data.

Meeting medical terminology needs--the Ontology-Enhanced Medical Concept Mapper: This paper describes the development and testing of the Medical Concept Mapper, a tool designed to facilitate access to online medical information sources by providing users with appropriate medical search terms for their personal queries. Our system is valuable for patients whose knowledge of medical vocabularies is inadequate to find the desired information, and for medical experts who search for information outside their field of expertise. The Medical Concept Mapper maps synonyms and semantically related concepts to a user's query. The system is unique because it integrates our natural language processing tool, i.e., the Arizona (AZ) Noun Phraser, with human-created ontologies, the Unified Medical Language System (UMLS) and WordNet, and our computer generated Concept Space, into one system. Our unique contribution results from combining the UMLS Semantic Net with Concept Space in our deep semantic parsing (DSP) algorithm. This algorithm establishes a medical query context based on the UMLS Semantic Net, which allows Concept Space terms to be filtered so as to isolate related terms relevant to the query. We performed two user studies in which Medical Concept Mapper terms were compared against human experts' terms. We conclude that the AZ Noun Phraser is well suited to extract medical phrases from user queries, that WordNet is not well suited to provide strictly medical synonyms, that the UMLS Metathesaurus is well suited to provide medical synonyms, and that Concept Space is well suited to provide related medical terms, especially when these terms are limited by our DSP algorithm.

Exploiting medical hierarchies for concept-based information retrieval: Search technologies are critical to enable clinical staff to rapidly and effectively access patient information contained in free-text medical records. Medical search is challenging as terms in the query are often general but those in relevant documents are very specific, leading to granularity mismatch. In this paper we propose to tackle granularity mismatch by exploiting subsumption relationships defined in formal medical domain knowledge resources. In symbolic reasoning, a subsumption (or 'is-a') relationship is a parent-child relationship where one concept is a subset of another concept. Subsumed concepts are included in the retrieval function. In addition, we investigate a number of initial methods for combining weights of query concepts and those of subsumed concepts. Subsumption relationships were found to provide strong indication of relevant information; their inclusion in retrieval functions yields performance improvements. This result motivates the development of formal models of relationships between medical concepts for retrieval purposes.

Combining Bayesian Text Classification and Shrinkage to Automate Healthcare Coding: A Data Quality Analysis. This article analyzes the data quality issues that emerge when training a shrinkage-based classifier with noisy data. A statistical text analysis technique based on a shrinkage-based variation of multinomial naive Bayes is applied to a set of free-text discharge diagnoses occurring in a number of hospitalizations. All of these diagnoses were previously coded according to the Spanish Edition of ICD9-CM. We deal with the issue of analyzing the predictive power and robustness of the statistical machine learning algorithm proposed for ICD-9-CM classification. We explore the effect of training the models using both clean and noisy data. In particular our work investigates the extent to which errors in free-text diagnoses propagate to the classification model. A measure of predictive accuracy is calculated for the text classification algorithm under analysis. Subsequently, the quality of the sample data is incrementally deteriorated by simulating errors in the text and/or codes. The predictive accuracy is recomputed for each of the noisy samples for comparison purposes. Our research shows that the shrinkage-based classifier is a valid alternative to automate ICD9-CM coding even under circumstances in which the quality of the training data is in question.

Large Scale Diagnostic Code Classification for Medical Patient Records: The coding approach currently used in hospitals relies heavily on manual labeling performed by skilled and/or not so skilled personnel. This is a very time consuming process, where the person involved reads the patient chart and assigns the appropriate codes. Moreover, this approach is very error prone given the huge number of CPT and ICD9 codes. A recent study (Benesch et al., 1997) suggests that only 60%-80% of the assigned ICD9 codes re- flect the exact patient medical diagnosis. This can be partly explained by the fact that coding is done by medical abstractors who often lack the medical expertise to properly reach a diagnosis. Two situations are prevalent: ”over-coding”. Both situations translate into significant financial loses: for insurance companies in the first case and for hospitals in the second case. Additionally, accurate coding is extremely important because ICD9 codes are widely used in determining patient eligibility for clinical trials as well as in quantifying hospital compliance with quality initiatives A critical, yet not very well studied problem in medical applications is the issue of accurately labeling patient records according to diagnoses and procedures that patients have undergone. This labeling problem, known as coding, consists of assigning standard medical codes (ICD9 and CPT) to patient records. Each patient record can have several corresponding labels/codes, many of which are correlated to specific diseases. The current, most frequent coding approach involves manual labeling, which requires considerable human effort and is cumbersome for large patient databases. In this paper we view medical coding as a multi-label classification problem, where we treat each code as a label for patient records. Due to government regulations concerning patient medical data, previous studies in automatic coding have been quite limited. In this paper, we compare two efficient algorithms for diagnosis coding on a large patient dataset.

3. EXISTING SYSTEM

Most of the current health providers organize and code the medical records manually. This workflow is extremely expensive because only well-trained experts are properly competent for the task. Therefore, there is a growing interest to develop automated approaches for medical terminology assignment. The presented techniques can be classified into two categories: rule-based and machine learning approaches. Rule-based approaches play a principle role in medical terms assignments. They generally discover and construct effective rules by making strong uses of the semantic, syntactic, morphological and pragmatic feature of natural language. It has been found that these methods have significant positive effects on the real systems. Machine learning approaches build inference models from medical data with known annotations and then apply the trained models to unseen data for terms prediction. A multi-label large-margin formulation that clearly incorporated the inter-terminology structure and prior domain knowledge concurrently. This approach is possible for small terminology set but is questionable in real-life settings where thousands of terminologies need to be considered.

In existing model, when the first classifier made a known error, the output of the second classifier was used instead to give the final prediction.

- The existing system use the local mining and global learning.
- Local mining aims to locally code the medical records by extracting the medical concepts.
- The global learning analysis collaboratively learns missing key concepts

Disadvantages:

- Information loss occurs.
- Low precision for data records.
- Difficult to analysis irrelevant medical concepts.

4. PROPOSED SYSTEM

We propose a novel scheme that is able to code the medical records with corpus-aware terminologies. The planned method consists of two equally opposed to components, namely, local mining and global learning. Local mining aims to locally code the medical records by extracting the medical concepts from individual record and then mapping them to terminologies based on the external authenticated vocabularies. We establish a tri-stage framework to accomplish this task, which includes noun phrase extraction, medical concept detection and medical concept normalization. As a byproduct, a corpus-aware terminology vocabulary is naturally construct, which can be utilize as terminology space for further learning in the second component. However, local mining approach may suffer from the problem of information loss and low precision due to the possible lack of some key medical concepts in the medical records and the presence of some irrelevant medical concepts. We thus propose global learning to complement the local medical coding in a graph-based approach. It collaboratively learns missing key concepts and propagates precise terminologies among underlying connected records over a large collection. Besides the semantic similarity among medical records and terminology-sharing network, the inter-terminology and inter-expert relationships are seamlessly integrated in the proposed model.

Advantages:

Ontology based similarity measure has some advantages over other measures.

- The proposed system uses the ontology.
- It used for find the inter-terminology and inter-expert relationships.
- The inter expert relationships are inferred from the experts' historical data.
- Overcome the mismatch problems of data.
- Decrease the unrelated sibling terminologies.
- Remove data Inconsistency

5. CONCLUSION AND FUTURE WORK

This paper presents a medical terminology assignment scheme to bridge the vocabulary gap between health seekers and healthcare information. The system comprises of two mechanism, local mining and global learning. The previous create a tri-stage framework to locally code every medical record. Though, the local mining approach might suffer from information loss and low precision, which are caused by the absence of key medical concepts and the presence of the irrelevant medical concepts. This motivates us to propose a global learning approach to compensate for the insufficiency of local coding approach. The second component collaboratively learns and propagates terminologies among underlying connected medical records. It enables the integration of heterogeneous information. To flexibly organize the unstructured medical content into user needs-aware ontology by leveraging the recommended medical terminologies.

6. FUTURE WORK

The future work based on the idea of preprocessing both the text and pattern strings as against to other existing algorithms which either pre process text or pattern or does no preprocessing such as Brute Force algorithm. The behavior of the algorithm depends on the minimum occurrence character in the pattern. The search is performed only in the segments where the minimum character of the pattern is found thus skipping the comparisons in the segments not containing the same which reduces the number of comparisons being performed.

REFERENCES

[1] L. Nie, T. Li, M. Akbari, and T.-S. Chua, "Wenzher: Comprehensive vertical search for healthcare domain," in Proc. Int. ACM SIGIR Conf., 2014, pp. 1245–1246.

[2] AHIMA e-HIM Work Group on Computer-Assisted Coding, "Delving into computer-assisted coding," J. AHIMA, vol. 75, pp. 48A–48H, 2004.

[3] G. Leroy and H. Chen, "Meeting medical terminology needs-the ontology-enhanced medical concept mapper," IEEE Trans. Inf. Technol. Biomed., vol. 5, no. 4, pp. 261–270, Dec. 2001.

[4] G. Zuccon, B. Koopman, A. Nguyen, D. Vickers, and L. Butt, "Exploiting medical hierarchies for concept-based information retrieval," in Proc. Australasian Document Comput. Symp., 2012, pp. 111–114.

[5] E. J. M. Laur_ıa and A. D. March, "Combining Bayesian text classification and shrinkage to automate healthcare coding: A data quality analysis," J. Data Inf. Quart., vol. 2, no. 3, p. 13, 2011.

[6] L. Yves A., S. Lyudmila, and F. Carol, "Automating ICD-9-cm encoding using medical language processing: A feasibility study," in Proc. AMIA Annu. Symp., 2000, p. 1072.

[7] C. Dozier, R. Kondadadi, K. Al-Kofahi, M. Chaudhary, and X. Guo, "Fast tagging of medical terms in legal text," in Proc. Int. Conf. Artif. Intell. Law, 2007, pp. 253–260.

[8] L. V. Lita, S. Yu, S. Niculescu, and J. Bi, "Large scale diagnostic code classification for medical patient records," in Proc. Conf. Artif. Intell. Med., 1995.

[9] L. S. Larkey and W. B. Croft, "Automatic assignment of icd9 codes to discharge summaries," PhD dissertation, Dept. Comput. Sci., Univ. Massachusetts at Amherst, Amherst, MA, USA, 1995.