BRIDGING THE LEXICON GAP AMONG MEDICAL TERMINOLOGY

K.Vinothini
N.Venkatesan

PG Scholar ,Department of Computer Science and Engineering ,Sri Vidya College of Engineering & Technology,Virudhunagar
Assistant professor ,Department of Computer Science and Engineering, Sri Vidya College of Engineering & Technology,Virudhunagar

Abstarct

To develop the interaction and to compete the level of gap between different medical terminology and the common people in need of health details. This paper presents the novel scheme to code the medical record using the local mining and global learning terminology. In the local mining approach will face the problem of lower precision due to the unavailability of key medical concepts and irrelevant data on medical terminology. On the other side the global learning approach will overcome the drawbacks in local mining approach. The coordination of systematized nomenclature of medicine clinical terminology and International classification of diseases methodology is used in the dictionary comparison instead of systematized nomenclature of medicine clinical terminology alone. Global mining approach is done by finding the missing key terminology and data loss in medical concepts is avoided by analysing the social neighbours. The performance of our approach is compared with the existing techniques. The results show that our integrated approach significantly improvesnthe performance of web database systems and outperforms its counterparts

Keywords *:local mining, parts of speech, noun extraction, global learning .*

I.INTRODUCTION

Information technologies are transforming the ways healthcare services are delivered, from patients passively technologies are transforming the ways taking up their doctors' orders to patients keenly looking for online information that concerns their health for both professionals and health seekers this forum will be more attractive. For professionals, this will improve their reputations different their colleagues and patients, support their practical knowledge from interactions with other doctors in addition to attract more new patients. For patients, this system provides trusted answers especially fo

complicated problems. a incredible number of medical records have been accumulated in their repositories, mostly users might directly find good answers by searching from the record archives, rather waiting for the experts responses. In many cases, the community generated content, is not directly used due to the vocabulary gap. Users with diverse backgrounds do not share the same vocabulary. This focused on hospital generated health data or health provider released sources by using the is approaches.

loosely coupled rule-based and machine learning By comparing these data, the emerging community generated health data is more informal, in terms of inconsistency, complexity and ambiguity, which face the challenges for data access and analytics. Most of the previous work simply utilizes the external medical dictionary to code the medical records rather than considering the corpus-aware terminologies. Their dependence on the independent external knowledge might bring in inappropriate terminologies. Constructing a corpus-aware terminology vocabulary to trim the unrelated terminologies of exact dataset and limit the candidates is the tough issue we are facing. In addition, the variety of heterogeneous cues was often not sufficiently exploited simultaneously. So, a robust integrated framework to draw the strength from various resources and models is still expected. Though, local mining approach might 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.

Thus we propose global learning to complement the local medical coding in a graph-based approach. It collaboratively learns missing key concepts and propagates accurate terminologies different underlying connected records over a large collection. As well the semantic similarity different medical records and terminology-sharing network, the inter-terminology and inter-expert relationships are flawlessly integrated in the proposed model. The inter-terminology relationships are mined by exploiting the external clear ontology, which are able to alleviate the granularity mismatch problems and reduce the irrelevant sibling terminologies. The inter expert relationships are inferred from the expert's chronological data.

II. RELATED WORK

Machine Learning

Emerging technologies have brought in extreme changes in medicine, in diagnosis in addition to medical assistance. With the active online trends, many forums have been developed with provides the seeker with instant medical advice or recommendations. Example of one such forum would be WebMD. Such forums favour both the experts/professional and the user looking for information. In case of the professionals, it provides a platform to interact with renowned medical experts, strengthen their knowledge, and attract new patients. Also such forums provide an opportunity to increase their reputation different their colleagues and patients. For users, these give up instant answers to their queries and from trusted sources. These forums rely on community related data rather than waiting for the expert's response or browsing to see various related information from the web and picking the accurate information from relevant information, thus relying on community generated data. Though it might not be advisable to directly use the community generated data as it might result in a vocabulary gap, where users from different backgrounds might not be able to understand the various medical terms that has been used as that might not use the same vocabulary.

Ranking Based Answer Generation

As the emerging trends and communication technologies are developed, there is an alternative to obtain information online, owning to the following facts. First, information seekers are able to post their exact questions on any topic and obtain answers provided by other participants. Patients are able to get better answers than simply using search engines. Second, in comparison with programmed QA systems usually get the answers with better quality as they are generated based on human aptitude. The generated answers will be better than the answers searched by the browsers and accurate result is obtained. Third, over times, a tremendous number of QA pairs have been accumulated in their repositories, and it facilitates the preservation and search of answered questions. Existing forums mostly support only textual answers. Unfortunately, textual answers might not provide a good result. So, the textual answers in cQA can be significantly enhanced by adding multimedia contents, and it will provide answer seekers more comprehensive information and better experience

Recommendation Based On Feedback

Significant advancements made in the field of recommender systems, news recommendation is still for recommendation. The algorithm driving most of the commercially popular recommender systems has been mutual filtering. While collaborative filtering works exceptionally well when the number of items and users are fixed, it starts to fail when they are not. Especially, in the news domain where the life time of a news story is in general and the number of stories and their content is updated. This makes the problem of recommending relevant news articles extremely challenging. Moreover, a news recommender systems need to cater to factors like freshness and dynamic popularity of the articles. Added to the above concerns is the reality check that the news needs to be personalized which requires understanding the user's chronological consumption behavior and other localized factors

III System Architecture

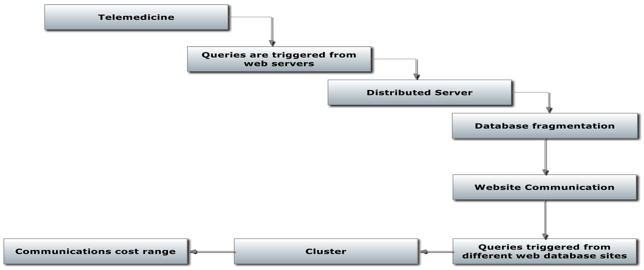

Telemedicine
Queries are triggered from web servers
Distributed Server
Database fragmentation
Website Communication
Queries triggered from different web database sites
Cluster
Communications cost range

Figure 1

Local Mining:

Medical concepts are defined as medical domain-specific noun phrases, and medical terminologies are referred to as authenticated phrases by well- known organizations that are used to accurately describe the human body and associated components, conditions and processes in a science -based manner .

To accomplish this task, we establish a tri –stage framework. Specifically, given a medical record, we first extract the embedded noun phrases.

We then identify the medical concepts from these noun phrases by measuring their specificity. Finally, we normalize the detected medical concepts to terminologies.

Noun Phrase Extraction

To extract all the noun phrases, we initially assign part-ofspeech tags to each word in the given medical record by Stanford POS tagger. We then pull out sequences that match a fixed pattern as noun phrases

(Adjective|Noun)*(Noun preposition)?(Adjective|Noun)*Noun

Medical Concept Detection

In this stage we use concept entropy impurity to measure the domain relevance concept.

CEI(C)=-P(Di|c)logP(Di|c)

Where D represents our medical corpus and a general domain corpus and P(Di|c) denotes the probability that a concept is related to a specified domain Di;

Medical concept normalization

In the medical concept normalization the words are compared with the medical terminology in the dictionary meant for it. There exist numerous authenticated vocabulary which includes ICD,UMLS and SNOMED CT .In this paper we use the coordination of SNOMED CT and ICD 10.

Token Matching Algorithm

```
Step 1
for each word in list:
add entry to the Matching Matrix
for new column:
Intersect new word with
cell from matching table
Sort the matching array in descending order based off the
scores
for each row in the matrix:
start at the right most cell
STEP 2
if the top score for the cell is 1.0
add cell details to current best match list, update
current match score.
recursively call STEP 3 on cell (row=column+2,
column=right)
```

```
else:
move one column left to the next cell
or
the right-most cell of the next row if left cell
empty
repeat STEP 3 until visited all cells
FINISH
return match
```

SNOMED and ICD Work Together

Through mapping and integration, SNOMED-CT is linked with other classifications or terminologies so that:

- Healthcare data collected for one purpose can be used for another purpose
- Data can be more easily migrated to newer database schemas and formats
- avoiding multiple data entry and reducing the risk of higher cost and errors Data can be entered once and reused.

Clinical data captured at the point of care can be effectively and efficiently used for administrative purposes such as vital and health statistics trending, compensation and health policy decision-making The use of a map from SNOMED-CT to ICD-10-CM and ICD-10-PCS will allow clinical information captured at a very granular level to be aggregated for reporting and statistical analysis purposes. Mapping a reference terminology to modern classification systems:

Mapping between SNOMED-CT and ICD is an imperfect science. For a computer's purposes it is very difficult to adequately represent some of the ICD coding conventions. The codes produced by the crossmap will need to be evaluated in the applicable reporting rules and context of the complete medical record and compensation requirements before being submitted to other external entities and payers . The construct, representation of modern clinical medicine and greater specificity provided by ICD-10-CM and ICD-10-PCS will greatly facilitate the computerized generation of ICD codes ultimately anticipated in an EHR environment.*ICD-9-CM,ICD-10-CM,andICD-10-PCS*.ICD-9-CM, ICD-10-CM, and ICD-10-PCS are referred to as "classifications." They are also sometimes called "administrative terminologies" because they are commonly used for other administrative purposes external reporting , such as epidemiological analyses and statistical , compensation for healthcare services, and public health reporting.

Mapping: Creating Connections

Coding and mapping are very different activities. Coding involves the use of and other clinical data contained in an individual patient health record and clinician documentation as the source for determining the appropriate code assignment within a terminology or classification. Coding conventions and guidelines are applied in determining code assignment. Also, appropriate code selection sometimes depends upon the context of a specific patient record .In order to accurately and fully represent information in a health record we have to add additional information to a concept, and this information fundamentally changes the meaning of the concept, it is called "context."

Mapping is the process of linking content from one terminology to a classification or to another. Maps result in an expression of the relationships between the terminologies or classification systems involved. Mapping requires deciding how concepts in different terminologies are similar,match or differ. It provides a link between terminologies and classifications in order to:

- Use data collected for one purpose for another purpose
- Retain to newer database formats and schemas when migrating the value of data
- Avoid the associated risk of increased cost and errors and entering data many times

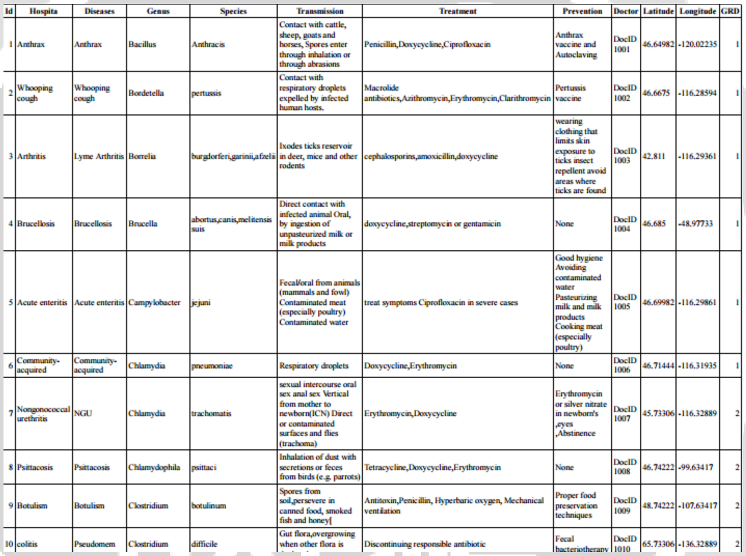
Unlike coding, mapping is not specific to a particular patient encounter. Context is not available as part of the mapping process. map creation generally involves an computerized translation software engine. Computerized maps create efficiency by patient data integration across a wide variety of applications and minimizing duplicative data entry.

Maps regulate linkages to a certain extent and therefore improve coding precision simply and efficiently through computerized algorithms. Mapping considers different purposes, levels of detail, and coding guidelines of source (terminology being mapped from) and target (terminology being mapped to).

The mapping process employs a standard method in which the classification description principles terminology or heuristics context are interpreted between systems. It begins with the development of (rules of thumb used for solving problems) heuristics and guidelines that support the use case or purpose of the map, respecting the conventions of the target and source to preserve the flexibility and granularity of both. Defined mapping rules must be consistently developed is applied without compromising clinical integrity to minimize incompatibilities.

Because terminologies have different intended uses and structures, crossmapping does not necessarily involve one-to-one relationships. There can also be many-to-one and one-to-many relationships, as well as concepts that are not mappable because the concept only exists in the target or source terminology.

Dataset



Id	Hospita	Diseases	Genus	Species	Transmission	Treatment	Prevention	Doctor	Latitude	Longitude	GRD
1	Anthrax	Anthrax	Bacillus	Anthracis	Contact with cattle, sheep, goats and horses, Spores enter through inhalation or through abrasions	Penicillin,Doxycycline,Ciprofloxacin	Anthrax vaccine and Autoclaving	DocID 1001	46.64982	-120.02235	1
2	Whooping cough	Whooping cough	Bordetella	pertussis	Contact with respiratory droplets expelled by infected human hosts.	Macrolide antibiotics,Azithromycin,Erythromycin,Clarithromycin	Pertussis vaccine	DocID 1002	46.6675	-116.28594	1
3	Arthritis	Lyme Arthritis	Borrelia	burgdorferi,garinii,afzelii	Ixodes ticks reservoir in deer, mice and other rodents	cephalosporins,amoxicillin,doxycycline	wearing clothing that limits skin exposure to ticks insect repellent avoid areas where ticks are found	DocID 1003	42.811	-116.29361	1
4	Brucellosis	Brucellosis	Brucella	abortus,canis,melitensis suis	Direct contact with infected animal Oral, by ingestion of unpasteurized milk or milk products	doxycycline,streptomycin or gentamicin	None	DocID 1004	46.685	-48.97733	1
5	Acute enteritis	Acute enteritis	Campylobacter	jejuni	Fecal/oral from animals (mammals and fowl) Contaminated meat (especially poultry) Contaminated water	treat symptoms Ciprofloxacin in severe cases	Good hygiene Avoiding contaminated water Pasteurizing milk and milk products Cooking meat (especially poultry)	DocID 1005	46.69982	-116.29861	1
6	Community-acquired	Community-acquired	Chlamydia	pneumoniae	Respiratory droplets	Doxycycline,Erythromycin	None	DocID 1006	46.71444	-116.31935	1
7	Nongonococcal urethritis	NGU	Chlamydia	trachomatis	sexual intercourse oral sex anal sex Vertical from mother to newborn(ICN) Direct or contaminated surfaces and flies (trachoma)	Erythromycin,Doxycycline	Erythromycin or silver nitrate in newborn's ,eyes ,Abstinence	DocID 1007	45.73306	-116.32889	2
8	Psittacosis	Psittacosis	Chlamydophila	psittaci	Inhalation of dust with secretions or feces from birds (e.g. parrots)	Tetracycline,Doxycycline,Erythromycin	None	DocID 1008	46.74222	-99.63417	2
9	Botulism	Botulism	Clostridium	botulinum	Spores from soil,persevere in canned food, smoked fish and honey[	Antitoxin,Penicillin, Hyperbaric oxygen, Mechanical ventilation	Proper food preservation techniques	DocID 1009	48.74222	-107.63417	2
10	colitis	Pseudomem	Clostridium	difficile	Gut flora,overgrowing when other flora is	Discontinuing responsible antibiotic	Fecal bacteriotherapy	DocID 1010	65.73306	-136.32889	2

Figure 2

Global Learning

In this approach the incompleteness and lower precision are overcome by using the relationship identification. Inter-terminology relationship and Inter –Expert Relationship

The info loss is caused by some lost key ideas of the given medical record. They, however, are probably present in the semantically similar neighbours.

The global learning is able to discover the missing key concepts from underlying connected medical data and strongly link them to the given medical record.

To fairly evaluate our unsupervised learning approach, other supervised graph-based learning methods were not listed here.

For each method mentioned above, the involved parameters were carefully tuned, and the parameters with the best performances were used to report the final comparison results.

Website comparision which reduces the communication cost than before techniques used in the medical terminology retrival.

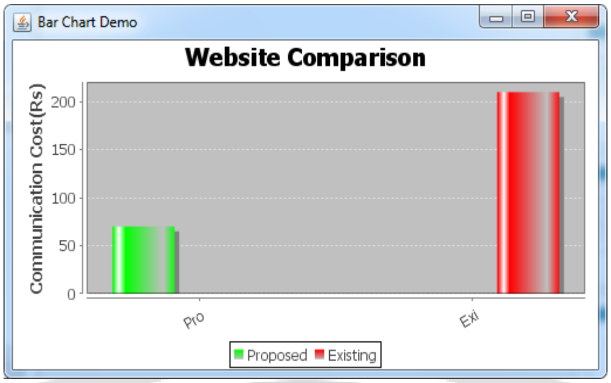


Figure 3

Cost comparison which show the difference in the cost for the retrival of the medical terminology concept

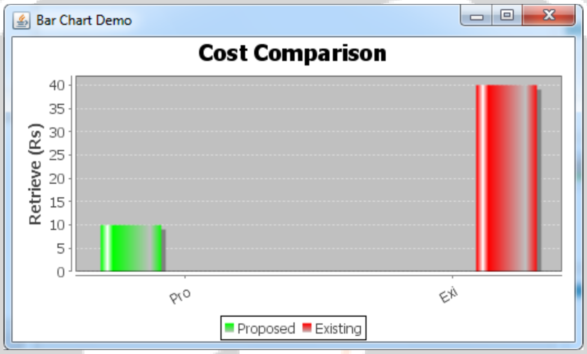


Figure 4

IV.Conclusion

In this work, we proposed a new approach to promote the retrieval performance. Our approach integrates three enhanced computing services' techniques namely, database fragmentation, network sites clustering and fragments allocation. This includes the SNOMED CT and ICD 10 in the database retrieval technique. We develop these techniques to solve technical challenges, like distributing data fragments among multiple web servers, handling failures, and making tradeoff between data availability and consistency. We propose an estimation model to compute communications cost which helps in finding cost-effective data allocation solutions. In the comparison of dictionary here we use the coordination of SNOMED CT and ICD 10. The novelty of our approach lies in the integration of web database sites clustering as a new component of the process of WTDS design in order to improve performance and satisfy a certain level of quality in web services. We perform both external and internal evaluation of our integrated approach. In the internal evaluation, we measure the impact of using our techniques on WTDS and web service performance measures like communications cost, response time and throughput. In the external evaluation, we compare the performance of our approach to that of other techniques in the literature. The results show that our integrated approach significantly improves services requirement satisfaction in web systems.

V.References

[1] AHIMA e-HIM Work Group on Computer-Assisted Coding,“ insant medical assistance through machine learning” j. ahIMA, vol. 75,pp. 48A–48H, 2015

[2] E. J. M. Laurıa and A. D. March, Beyond Text QA Multimedia diverse relevance ranking based Answer Generation by Extracting Web ,” J. Data Inf. Quart., vol. 2, no. 3, p. 13, 2014

[3] G. Leroy and H. Chen, “Personalized News Recommendation based On Implicit Feedback” IEEE Trans. Inf.Technol. Biomed., vol. 5, no. 4, pp. 261–270, Dec. 2014.

[4] L. Nie, T. Li, M. Akbari, and T.-S. Chua, “Multimedia Based Community Question Answer by HarvestingWeb Based Information,” in Proc. Int. ACMSIGIR Conf., 2014, pp. 1245–1246.

[5] G. Zuccon, B. Koopman, A. Nguyen, D. Vickers, and L. Butt,“ Learning to Recommend Descriptive Tags for Questions in Social Forums,” in Proc. Australasian Document Comput. Symp., 2013,pp. 111–114.

[6] . Nie, M. Akbari, T. Li, and T.-S.Chua, “A joint local-globalapproach for medical terminology assignment,” in Proc. Int. ACMSIGIR Conf., 2014.

[7] L. Yves A., S. Lyudmila, and F. Carol, “Ontology Based Medical Terminology System”in Proc. AMIA Annu. Symp., 2000, p. 1072.

[8] S. Hina, E. Atwell, and O. Johnson, “Semantic tagging of medicalnarratives with top level concepts from SNOMED CT healthcaredata standard,” Int. J. Intell.Comput.Res., vol. 2, pp. 204–210, 2010.

[9] J. Patrick, Y. Wang, and P. Budd, “An computerized system for conversion of clinical notes into snomed clinical terminology,” inProc. 5th Australasian Symp. ACSW Frontiers, 2007, pp. 219–226.

[10] L. S. Larkey and W. B. Croft, “Automatic assignment of icd9 codesto discharge summaries,” PhD dissertation, Dept. Comput.Sci.,Univ. Massachusetts at Amherst, Amherst, MA, USA, 1995.

[11] H. Suominen, F. Ginter, S. Pyysalo, A. Airola, T. Pahikkala, S.Salanter, and T. Salakoski, “Machine learning to automate theassignment of diagnosis codes to free-text radiology reports: Amethod description,” in Proc. ICML Workshop Mach. Learn.Health-Care Appl., 2008.

[12] L. V. Lita, S. Yu, S. Niculescu, and J. Bi, “Large scale diagnosticcode classification for medical patient records,” in Proc. Conf. Artif.Intell. Med., 1995.

[13] W. R. Hersh and H. David, “Information retrieval in medicine:Thesaphire experience,” J . Am e r.Soc .I n f .S c i ., vol. 46, no. 10,pp. 743–747, 1995.

[14] Q. Zhou, W. W. Chu, C. Morioka, G. H. Leazer, and H. Kangarloo,“Indexfinder: A method of extracting key concepts from clinicaltexts for indexing,” in Proc. AMIA Annu. Symp., 2003, pp. 763–767.

[15] Y. Wang and J. Patrick, “Mapping clinical notes to medical terminology at point of care,” in Proc. Workshop Current Trends Biomed.Natural Lang. Process., 2008, pp. 102–103.

[16] 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.