



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

A Survey on Associating the Lexicon Gap between Health Quester and Healthcare Lore

Vinothini.K¹, Venkatesan.N²

PG Scholar, Dept. of CSE, Sri Vidya College of Engineering & Technology, Virudhunagar, Tamil Nadu, India¹

Assistant Professor, Dept. of CSE, Sri Vidya College of Engg. & Tech, Virudhunagar, Tamil Nadu, India²

ABSTRACT: To develop the interaction and to complete 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. This is done by finding the missing key terminology and data loss in medical concepts is avoided by analysing the social neighbours.

KEYWORDS: Medical terminology, Lexicon Gap, Local Mining, Global Mining

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 for 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 isolated or loosely coupled rule-based and machine learning approaches.

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.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

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, owing 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. LITERATURE REVIEW

AUTHOR & YEAR	TITLE	METHODOLOGY	ADVANTAGES	DISADVANTAGES
Liqiang Nie Mohammad Akbari Tao Li 2014	A Joint Local- Global Approach for Medical Terminology Assignment	The system define medical concepts as medical domain-exact noun phrases, and medical terminologies 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.	It is worth mentioning that there already exist several efforts dedicated to research on automatically mapping medical records to terminologies	The varieties of heterogeneous cues were often not adequately exploited simultaneously



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

Guido Zuccon , Bevan Koopman , Anthony Ngyue , Deanne Vickers, Luke Butt 2012	Exploiting Medical Hierarchies for Concept-based Information Retrieval	In this paper they proposed to tackle granularity mismatch by exploiting relationships denned 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.	This result motivates the development of formal models of relationships between medical concepts for retrieval purposes	This often fail to identify medical entities that are referred to with different terms, such as the Synonymous terms `heart attack' and `myocardial disorder'.
Hanna Suominen Filip Ginter Sampo Pyysalo Antti Airola Tapio Pahikkala Sanna Salanter`a Tapio Salakoski 2008	Machine Learning to Automate the Assignment of Diagnosis Codes to Free-text Radiology Reports: a Method Description	The system introduces a multi-label classification sys-tem for the automated assignment of diagnostic codes to radiology reports. The system is a cascade of text enrichment, feature selection and two classifiers.	This study provides insight into the development of applications for real-life usage, which are currently rare	This would not further improve the performance.
Saman Hina Eric Atwell Owen Johnson 2014	Semantic Tagging of Medical Narratives with Top Level Concepts from SNOMED CT Healthcare Data Standard	In this paper they presented the ongoing research on SNOMED CT concept extraction from discharge summaries using natural language processing and introducing SNOMED CT core concepts as a single gazetteer list for concept extraction.	It is clear that SNOMED CT will be the most suitable choice in the UK NHS (National Health Service)	The problem with defined sections is that they are not always in one format and defined in different Cases.
Yan Yan Glenn Fung Jennifer G. Dy Romer Rosales 2010	Medical Coding Classification by Leveraging Inter- Code Relationships	In this paper, they introduced a multi-label large-margin classifier that automatically learns the underlying inter-code structure and allows the controlled incorporation of prior knowledge about medical code relationships	The proposed multi-label classifier outperforms related multi-label models in this problem	It do not take advantage of prior domain knowledge efficiently
Mengqiu Wang Wanxiang Che Christopher D. Manning 2010	Joint Word Alignment and Bilingual Named Entity Recognition Using Dual Decomposition	The system introduces additional cross-lingual edge factors that encourage agreements between tagging and alignment decisions. We design a dual decomposition inference algorithm to perform joint decoding over the combined alignment and NER output space.	This method give ups significant improvements in both NER and word alignment over state-of-the-art monolingual baselines.	Their entity tags might not agree due to inconsistency in annotation standards.
Mark Hirschkorn David Geelan 2010	Bridging the Research-Practice Gap: Research Translation and/or Research Transformation	The issue of the `research-practice gap' – the problematic relationship between research in education and educational practice – is one that has been widely reported in the literature. This critical literature review explores some of the causes and features of the gap, and suggests some possible approaches for addressing it	This is essentially a top-down model in which researchers and the knowledge they produce govern the content and practice of teacher preparation.	It will not all be focused on changes on the part of practitioners.
Yuchi Huang Qingshan Liu Shaoting Zhang	Image Retrieval via Probabilistic Hypergraph	In this paper, they proposed a new transductive learning framework for image retrieval, in which images are	The hypergraph ranking approach tends to assign the same label	It is not complete to represent the relations different images only



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

Dimitris N. Metaxas 2010	Ranking	taken as vertices in a weighted hypergraph and the task of image search is formulated as the problem of hypergraph ranking.	to vertices that share many incidental hyperedges	by pair wise simple graphs.
Yu-Ru Lin Jimeng Sun Paul Castro Ravi Konuru Hari Sundaram Aisling Kelliher	MetaFac: Community Discovery via Relational Hypergraph Factorization	In this paper they proposed MetaFac (MetaGraph Factorization), a framework that extracts community structures from various social contexts and interactions.	An efficient factorization method for community extraction on a given metagraph	It is a very strong assumption in social media since there might be events involving some but not all dimensions.

Table 1 literature study

IV. PROPOSED WORK

In this system process is only focused Question and Answers. This not keep up patient history .so we have implemented patient history maintenance process. The system proposes three-fold approach based on data fragmentation, database websites clustering and intelligent data distribution. This approach reduces the amount of data migrated between websites during execution; achieves cost-effective communications during processing and improves response time and throughput. This new approach intends to decrease data communication, increase system throughput, reliability, and data availability. The web sites are grouped into clusters by using our clustering service technique in a phase prior to data allocation. The purpose of this clustering is to reduce the communications cost needed for data allocation.

V. CONCLUSION AND FUTURE WORK

This paper presents a medical terminology assignment scheme to bridge the vocabulary gap between health seekers and healthcare knowledge. The scheme comprises of two components, local mining and global learning. The former establishes a tri-stage framework to locally code each 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 different underlying connected medical records. It enables the integration of heterogeneous information. Extensive evaluations on a real-world dataset demonstrate that our scheme is able to produce promising performance as compared to the prevailing coding methods. More importantly, the whole process of our approach is unsupervised and holds potential to handle large-scale data. In the future, we will investigate how to flexibly organize the unstructured medical content into user needs-aware ontology by leveraging the recommended medical terminologies.

REFERENCES

- [1] L. Nie, M. Akbari, T. Li, and T.-S. Chua, "A joint local-global approach for medical terminology assignment," in Proc. Int. ACMSIGIR Conf., 2014.
- [2] L. Nie, T. Li, M. Akbari, and T.-S. Chua, "Wenzher: Comprehensive vertical search for healthcare domain," in Proc. Int. ACMSIGIR Conf., 2014, pp. 1245–1246.
- [3] AHIMA e-HIM Work Group on Computer-Assisted Coding, "Delving into computer-assisted coding," J. AHIMA, vol. 75, pp. 48A–48H, 2004.
- [4] G. Leroy and H. Chen, "Meeting medical terminology needs-theontology-enhanced medical concept mapper," IEEE Trans. Inf. Technol. Biomed., vol. 5, no. 4, pp. 261–270, Dec. 2001.
- [5] 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.
- [6] E. J. M. Lauria 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.
- [7] 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.
- [8] 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.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

- [9] 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.
- [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] J. Patrick, Y. Wang, and P. Budd, "An automated system for conversion of clinical notes into snomed clinical terminology," inProc. 5th Australasian Symp. ACSW Frontiers, 2007, pp. 219–226.
- [13] W. R. Hersh and H. David, "Information retrieval in medicine:Thesaphire experience," J. Am e r. S o c . 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.