A Survey on Methods That Restrict The Consistency in Online Health Seeker -Clinician Intercourse

R. Meena Gomathi
PG Student
Department of Computer Science & Engineering
K.L.N.College of Engineering, Madurai, Tamilnadu, India

Dr. S. Miruna Joe Amali
Associate Professor
Department of Computer Science & Engineering
K.L.N.College of Engineering, Madurai, Tamilnadu ,India

Abstract

Online wellness programs have been assisting well-being condition monitoring, illness modelling and validation OF grounded medical treatment by medical text mining. There exists a discernible gap in community based health forums between the online health seekers and providers. A cavernous learning structure is used to infer the diseases given the queries of health seekers. This scheme has two key components. The first globally mines the discriminant medical signatures from raw features. The second estimates the raw features and their signatures as input nodes in one layer and hidden nodes in the subsequent layer. The inter-relations between these two layers are found. All-encompassing trials on a real-world dataset labelled by online doctors show the noteworthy performance gains.

Keywords: Online wellness program, Signatures, Medical Text Mining

I. INTRODUCTION

Medical text mining is an approach of exploring the hidden patterns in the data sets of the medical field. The predominant online health services offer an interactive platform, where health seekers can ask health-oriented questions while doctors provide the knowledgeable and trustworthy answers. The healthcare environment is information rich but knowledge poor. The volume of health centre summaries written in natural language is swiftly increasing; doctors need a device to routinely extract facts about diseases/treatments. Information technology is expedient for transmuting the health care evidence from patients to clinicians via question answering. Question answering (QA) is the method for automatically answering a question posed in natural language. The medical QA scheme is able to response medical questions according to universal question nomenclature. Compared to keyword-based search systems, it highly facilitates the communication amid human and computer by generally stating user's intention in plain sentences. QA's performance is hampered by complicated natural language processing (NLP) techniques. The system take input as, a natural language question and return a precise verses that provide the responses. Question answering forums attract the consideration of both patients and consultants. The patients are provided with a rapidly and trusted answer for simple as well as complex health concerns. The doctors are able to raise their reputation among their colleagues and patient, support their practical knowledge from interactions with other prominent consultants, as well as possibly attract more new patients. There is a lexis disparity among health seekers and health care provider, in order to amortize the gap local mining and global learning approaches are used. Generally, the community laid content, may not be directly available due to the medical verbose gap. Users do not share similar vocabulary. For e.g. Health Tap, is a question answering site for users to request fitness related queries. The queries are written by our own words. The similar question may be described in substantially multiple ways by two distinct health seekers of diverse background. The answer provided by the health care professionals may contain expression with different possible connotations, and non-standardized terms. The tags used generally may not be medical terminologies. For e.g., "heart attack" and "myocardial disorder" is similar medical terms referred by multiple experts. Users had encountered big challenges in reusing the archived files due to the illogicality between their quest terms and those gathered medical records. Automatically coding of medical records is immensely desired using standardized terminologies. This survey paper summarizes methods to normalize medical verbose in order to curtail the gap among health inquirer and provider.

II. LITERATURE REVIEW

L. Nie et al., [1] have presented their effort to inhibit the terminology gap among health inquirers and providers which deferred the cross system inter-operability. Most of the existing health providers organize and code the medical archives by hand. There is an emergent concern to progress automated approaches for medical jargon assignment. Local mining and global learning methods are conjointly exploited. Local mining targets to locally code the health registers by mining the medical concepts from discrete record and then mapping them to lexicons based on the exterior legitimate terminologies.

Health provider released sources by utilizing either isolated or loosely connects rule-based and machine learning approaches. Most of the ongoing health providers organize and code the medical records manually. This workflow is extremely expensive

because only well trained experts are properly capable for the task. Therefore, there is a growing interest to build up automated approaches for medical terminology assignment. The existing techniques can be considered into two categories: rule-based and machine learning methods.

A. Local Mining Approach

Medical concepts are referred to medical domain specific noun phrases and Medical terminologies are allude 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.

B. Global Learning Approach

Global learning is an important method, including local approach, and attempted to map the QA pairs directly to the entries in external dictionaries without any pruning. This method generally presents problems since the external dictionaries naturally cover relatively comprehensive terminologies and are far beyond the vocabulary scale of the given corpus. It may result in the deterioration in coding performance in conditions of efficiency and effectiveness. The problem is caused by the over-turned scope of Vocabularies, which may take in unpredictable noises and make the precise terminology selection challenging. As a result, a corpus aware vocabulary terminology is naturally constructed by local mining approach, which can be used as terminology gap for further learning.

M. Wang et al., [2] anticipated a methodology that inevitably regulates which sort of media data should be added for a textual response. It spontaneously gathers facts from the net to augment the answer. By processing an enormous set of QA duos and accumulating them to a group, this methodology can qualify a unique multimedia question answering (MMQA) approach as consumers can discover multimedia responses by equating their questions with those in the group. For a given QA pair, scheme proposed in [2] first predicts which type of medium is appropriate for enriching the original textual answer. Following that, it automatically generates a query based on the QA knowledge and then performs multimedia search with the query. Proposed diverse relevance ranking scheme for social image search, which is able to simultaneously take relevance and diversity into account. It leverages both visual information of images and the semantic information of tags.

Diverse relevance ranking (DRR) scheme for social image search is used. It is able to rank the images based on their relevance levels with respect to query tag while simultaneously considering the diversity of the ranking list. First, it estimates the relevance score of each image with respect to the query term as well as the semantic similarity of each image pair. The relevance estimation incorporates both the visual information of images and the semantic information of their associated tags into an optimization framework, and the semantic similarity is mined based on the associated tags of images. Estimated relevance scores and similarities, we then implement the DRR algorithm, which can be viewed as a greedy ordering algorithm that optimizes average diverse result. Finally, query-adaptive re-ranking and duplicate removal are performed to obtain a set of images and videos for presentation along with the original textual answer.

Berlanga, Rafae et al., [3] intended an approach to produce word-concept likelihoods from a Knowledge Base (KB) that aids as a foundation for numerous text mining jobs which not only takes into account the core patterns within the descriptions enclosed in the Knowledge Base (KB) but also those in texts presented from huge unlabeled corpora such as MEDLINE. This system attains a higher degree of precision than other state-of-the-art methods when estimated on the MSH WSD data set.

Ben Abacha et al., [4] proffered a practice that deals with different types of questions, including questions with more than one expected answer and more than one focus. It allows a deep examination for queries and corpora by means of different information extraction approaches. The scheme uses both cavernous scrutiny of query and documents in order to abstract info. Recognize and mine the health entities (e.g. diseases, drugs, symptoms). Difficultly level lies in the purpose of classifying semantic relations between these entities (e.g. treats, prevents, causes).

Frunza, Oana; Inkpen, Diana; Tran, Thomas [5] in the year 2011 has put forward a method to identify and disseminate healthcare info and understand the semantic relationships that occur amongst syndromes and cures. It involves two tasks (i) automatically recognizing sentences published in medical abstracts (Medline) which encompasses evidence about syndromes and cures, and relate to semantic relations that exist between diseases and treatments. (ii)The second assignment is dedicated on three semantic relations: Cure, Prevent, and Side Effect. The noun-phrases, verb-phrases, and biomedical concepts are recognised in the sentences. Unified Medical Language system (UMLS) idea representations are an information source which comprises a meta thesaurus, a semantic system, and the expert lexicon for biomedical domain.

F. Wang, N. Lee, J. Hu et al.,[6] proposed novel temporal event matrix demonstration and knowledge structure in aggregation with an in-depth justification on equally synthetic and real world datasets. This framework enables the depiction, abstraction, and mining of high order event structure and relationships within single and multiple event sequences. Temporal Event Signature Mining involves two processes, i) One-Sided Convolution ii) β-divergence. Diverse event sequences are mapped to a geometric image by training events as a structured spatial-temporal shape process. It optimizes the performance of large-scale incremental learning of group-specific temporal event signatures. Then validates the framework on synthetic data and on an electronic health record dataset.

III. RELATED WORK

L. Nie et al., [1] has used Concept Entropy Impurity (CEI) methodology to relatively detect and normalize the health concepts locally, and construct a corpus responsive vocabulary with the assistance of exterior understanding. The local mining consists of the noun phrase mining, medical ideas detection and normalization. The global mining comprises of inter expert association; inter terminology relationship, probabilistic hyper graph creation. Global learning model is fabricated to collaboratively augment the local coding outcomes. This model flawlessly assimilates numerous heterogeneous information cues.
Following graph shows performance analysis of the two methods, namely local and global mining.

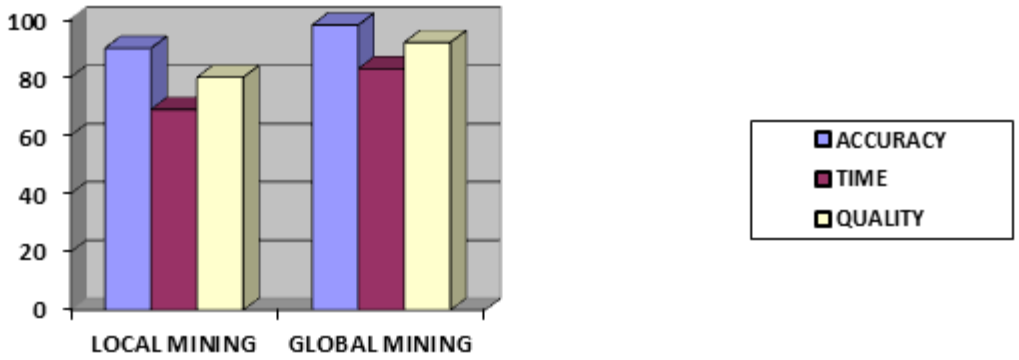

Fig. 1: analysis of the two methods, namely local and global mining

The methods proposed by L. Nie et al., [1] is completely indexed and thus the retrieval time is faster. In case of resource insufficiency the Query and the Question will be left in pending state till an expert arrives. When professionals go through the query, the responses not only dispatch to the wellness seekers and also update the local mining database for forthcoming instantaneous recovery to the related request from other users. The global learning approach is developed to compensate for the inadequacy of local coding method.

M. Wang et al., [2] has proposed three practices. For a question, retrieve question answer pair from the available question answering sites database dynamically and select an answer medium to enrich the textual answer. Then generate a query for the multimedia search, resulting data are undergoes duplicate elimination and irrelevant data removal by the help of graph based re ranking. Finally present the answer that contains textual data, images and videos. They 3 practices include

1) Answer medium selection- Given a QA pairs; it foretells whether the textual reply ought to be complemented with media evidence, and which kind of media files must be added.
2) Question generation for multimedia exploration-Given a QA pair, this component extracts three queries from the inquest, the response, and the QA duos, compatibly. The most informative question will be selected by a three-class ordering model.
3) Multimedia data selection and exhibition- Based on the produced queries, we gather image and video data with multimedia exploration engines. The gathering precisions for inquiry selection with diverse features such as POS histogram and search performance prediction (SPP) are used. By processing a large set of QA pairs and adding them to a pool, this approach can enable a novel multimedia question answering (MMQA) approach as users can find multimedia answers by matching their questions with those in the pool. The existing system lack of diversity of the generated media data.

Berlanga, Rafae et al., [3] has recommended a knowledge base which empowers huge quantity of machine learning approaches for diverse number of text mining tasks. Methods used are:

1) Word concept likelihoods estimate on WSD (Word Sense Disambiguation).
2) Document ranking- Derived from cross entropy amongst the word concepts and word concept model.

The biomedical knowledge base used in this model is UML'S which is a collection of massive amount of medical lexicons and ontology. Naive Bayes classifier turns as an orientation line to produce upper bound conclusions that are pre-eminent outfits for supervised data.

Frunza, Oana; Inkpen, Diana; Tran, Thomas [5] have suggested a mode to disseminate healthcare facts and understand the semantic associations that exist between diseases and treatments. As a sorting procedure, representative models such as: decision-based models probabilistic models (Naive Bayes (NB) and Complement Naive Bayes (CNB), adaptive learning (AdaBoost), a linear classifier (SVM) is utilized. This ML-based practice is used for constructing an application that is skilful in recognizing and splitting healthcare information swiftly. This obtains a reliable outcome that could be integrated in an application to be used in the health care domain.

Liqiang Nie et al.,[9] has put forward a model to analyse the user posted questions which excerpts corpus recognizant medical lexicons from those raw features. This step is termed to be called as "signature mining". A deep learning technique is used, which at preliminary level matches the symptoms discretely with a disease. It results in inferring various diseases for the

manifested signals. Correspondingly at descendant levels, combination of multiple symptoms is checked in order to predict the disease. This augments the level of accuracy in inferring the disease.

IV. RESULT ANALYSIS

The approach proposed by Liqiang Nie et al., [9] is showed in the table 1. Machine learning methods build inference models from medical data with well-known annotations and then apply the trained models to unseen data for terminology prediction. This paper benefits from the volume of unstructured community generated data and it is capable of handling various kinds of diseases effectively. It investigates and categorizes the information needs of health seekers in the community based health services and mines the signatures of their generated data. And the connected deep learning scheme that is able to infer the possible diseases given the questions of health seekers. This approach also permits unsupervised feature learning from other wide range of disease types. Therefore, it is generalizable and scalable.The performance of the developed system by using various classifiers is as follows:

Table – 1
Categorization of Performance based on Health Seeker Needs

Classifiers	*Questions*	*Questions+ Answers*	*Questions+ Answers+ Tags*
KNN	*87.4%*	*82.6%*	*83.4%*
SVM	*90.4%*	*83.6%*	*84.8%*
Naïve Bayes	*90.2*	*85.4%*	*84.4%*
Decision Tree	*89.4%*	*83.8%*	*85.2%*

V. CONCLUSION

This paper delivers an investigation of numerous methods used for categorizing medical relationships and limits the inequalities between online health seeker and provider. The advantages and shortcomings of each methodology are studied. The association between different tactics is quantified. It is reported that textual answers are preferred by online health seeker and the only hindrance is the medical verbose gap among health seeker and physician. This limitation is eradicated in paper [9] which deals with various medical text mining techniques to enrich replies in community based health care structures that shows a great performance improvement. The sstretched out assessments on a real world dataset validate that scheme mentioned in [9] is able to produce promising performance when equated to the prevalent coding procedures. The comprehensive approach is unsupervised and holds potential to handle large-scale data.

REFERENCES

[1] L. Nie, Y.-L. Zhao, M. Akbari, J. Shen, and T.-S. Chua, "Bridging the vocabulary gap between health seekers and healthcare knowledge," IEEE Trans. Knowl. Data Eng., vol. 27, no. 2, pp. 396–409, Jun. 2014.
[2] L. Nie, M. Wang, Y. Gao, Z.-J. Zha, and T.-S. Chua, "Beyond text qa: Multimedia answer generation by harvesting web information," IEEE Trans. Multimedia, vol. 15, no. 2, pp. 426–441, Feb. 2013.
[3] Jimeno Yepes, Antonio; Berlanga, Rafae ,"Knowledge based word-concept model estimation and refinement for biomedical text mining," Journal of Biomedical Informatics (Elsevier) ,vol.53,pp. 300-307,2015.
[4] Ben Abacha, Asma; Zweigenbaum, Pierre "MEANS: A medical question-answering system combining NLP techniques and semantic Web technologies," Journal of Information Processing & Management (Elsevier), vol.51,no. 5,pp- 570-594 ,2015.
[5] Frunza, Oana; Inkpen, Diana; Tran, Thomas," A Machine Learning Approach for Identifying Disease-Treatment Relations in Short Texts," IEEE Transactions on Knowledge and Data Engineering; Tran, vol.23, no.6, pp. 801-814, 2011.
[6] F. Wang, N. Lee, J. Hu, J. Sun, S. Ebadollahi, and A. Laine, "A framework for mining signatures from event sequences and its applications in healthcare data," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 2, pp. 272–285, Feb. 2013
[7] Fleuren, Wilco W.M.; Alkema, Wynand, "Application of text mining in the biomedical domain," Journal of Methods (Elsevier), vol.74, pp. 97-106, 2015.
[8] Zhang, Shaodian; Elhadad, Noémie, "Unsupervised biomedical named entity recognition: Experiments with clinical and biological texts," Journal of Biomedical Informatics (Elsevier) vol. 46, no. 6, pp. 1088-1098, 2013.
[9] Liqiang Nie, Meng Wang, Luming Zhang, Shuicheng Yan, Member, Bo Zhang and Tat-Seng Chua , "Disease Inference from Health-Related Questions via Sparse Deep Learning," IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 8, 2107-2119, august 2015.