www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 5 Issues 8 Aug 2016, Page No. 17530-17534

Disease Identification using Concept and Feature Relationship Analysis for Remote Health Services

[1] Dr. C. Kalaiselvi, Ph.D., [2] Ms. P. Krishnapriya,

Associate professor and Head,

Dept of Computer Applications,

Tiruppur Kumaran College for Women, Tirupur, Tamilnadu

Research scholar (M.Phil),

Dept of Computer Science,

Tiruppur Kumaran College for Women, Tirupur, Tamilnadu

ABSTRACT

Internet provides various resources for Remote health services. Web based communication models are adapted to manage the interaction between the remote people with medical experts. Remote health services are provided in two ways. They are reputable portals and community based health services. Reputable portal provides information related to the health care domains. Community based health services are built to support in a particular way based on health care solutions for the patients.

Community based health service models are applied to support disease identification process. Vocabulary and medical terminology are provided for the Patient and Doctor Communication. Medical concepts and diagnosis samples are required for the health services. Disease diagnosis is carried out with the question and answer communication data values. Disease identification is achieved using the sparse deep learning method. Local learning and global learning methods are adapted to analyse the raw features for signature identification process. Disease inferences are discovered using the sparse learning method.

Sparse learning method is enhanced to identify the discriminatory features. Medical term relationships are discovered with the medical domain based Ontology. Decision making is carried out with the concept relationship measurements. Symptoms and their importance are also considered in the decision support process.

Keywords: Decision Support System, Medical Gerontology, Remote Health Services, Semantic Analysis, Sparse Deep Learning.

1. INTRODUCTION

Due to the fast growth,of posted questions in CQA services, users may not be get their posted questions resolved in a short period. We randomly sampled 3, 640 questions from one popular CQA service Yahoo! Answers and kept track of the status of the questions. 640 questions were sampled from 26th first-level categories of Yahoo! Answers, with 140 questions in each division. After a day, we observed that only 434 (11.95%) of questions got answered and 726 (19.95%) questions in total will get resolved in two days. This finding shows a large number of posted questions cannot get changed in a short period. Same problem was also found in last research works. As a result, some users may not post new questions but they can reply on other questions to find information, if they cannot get their questions solved during a reasonable time period.

Due to the loss of an active question routing mechanism, a user is easily overwhelmed by the large number of open questions and cannot easily find questions if he/she is interested in answering even if he/she is willing to contribute his/her knowledge. Thus, there is a important gap between the current open questions and potential answers. To see through the gap, we present a new approach to Question Routing, which aims at routing open questions to suitable CQA users who may answer these questions. Question routing bluster several benefits. From the seeker's perspective, it can reduce the lagging of time between the time a question is posted and the time it is answered and it can potentially increase the user's satisfaction to CQA services. In return, the user may be more willing to contribute knowledge to the CQA service in the future. From providing the answer is important, because he/she will receive questions he/she is interested in instead of a large

number of unfiltered questions, the answers would become more enthusiastic in providing answers. From the CQA systems promote, by linking open questions with suitable answers, the CQA system could fully leverage users' answering affectionately, leading to the improvement of the CQA system, as well as the boost of the user's grip and loyalty to the system. In addition, to that CQA services share a lot of properties with other social media systems, the knowledge unwind can be ultimately applied to improving the performance of other social media applications.

2. RELATED WORK

Most of the current health providers arrange and insert the medical records manually. This workflow is highly expensive because only domain experts are properly capable for the task. Therefore, there is a growing interest to develop automatically approaches for medical terminology assignment. The remaining techniques can be categorized into two categories such as Machine learning and rule-based approaches. Rule-based approaches play a major role in medical terminology assignments. They generally discover and construct effective rules by making strong uses of the acceptable, syntactic, semantic and logical aspects of natural language. It has been found that these methods have significant positive effects on the real systems. Back in 1995, Hersh and David designed and developed a system name SAPPHIRE, which automatically assigned UM LS 5 rules to medical documents using a simple oral approach. Around one decade later, a system named Index Finder, proposed a new algorithm for generating all valid UM LS terminologies by permitting the set of words in the given text and then filtering out the insignificant concepts via syntactic and semantic filtering. Several efforts have developed automatically to convert free medical texts into medical terminologies by combining several common language processing methods, such as deriving, morphological analysis, lexicon augmentation, term composition and negation detection. These methods are entirely applicable to well-constructed discourses. It united UM LS, Word Net as well as Noun Phrase to capture the correct meaning of the queries.

In the year, support vector machine (SVM) and Bayesian ridge regression were first established on large-scale data set and obtained promising performance. Following that, a hierarchical model was studied exploited the structure of ICD-9 code set and demonstrated that their approach well performed the algorithms based on the classic vector space model. About ten years later, Sub continental introduced a precipitation of two classifiers to assign diagnostic terminologies to audio reports were obtained. In their model, when the first user made a known error, the outcome of the second classifier was used instead of giving the final result. Yenta. Proposed a multi-label large-margin formulation that accurate incorporated the inter-rule structure and above mentioned domain knowledge together. This approach is profitable for small rule but it is asked in real-life settings where thousands of rules need to be considered. Similar to our scheme Pakhomovetal tried to improve the coding performance by combining theadvantages of the Rule-based and Machine learning approaches.

Beyond medical domain, several prior efforts of corpus alignment and legal gap have been dedicated to other verticals. Chantal derived an integrated model that jointly aligns bilingual named entities between Chinese and English news. The work in the third bridged the management research-practice gap by describing their experiences with the network for business sustainability. A game platform was designed in fourth and was demonstrated how to enhance the inter-generation cultural communication in a family. These diverse efforts are all heuristic. Their rules and patterns are domain specific and cannot be generalized to other areas. Another example, the music semantic gap between textual query and audio content was remedied by annotation with fifth concepts. This approach can hardly be applied to medical terminology assignment directly due to the differences in models and content structures. It aims at marking music entities with common noun and adjective phrases, while our approach focuses on terminologies only.

3. HEALTH SEEKER CATEGORIES

Three different groups of Internet users emerge as worthy health seekers: those who used to look for health information on behalf of others; those with disabilities, those who care for others full time. A significant 57% of health seekers said that, last time they did a health search; they waiting for the information of others. This group is highly organized by male, female, healthy people and the middle aged between 30 and 49 years old. Parents are more similar than non-parents to have dedicated their last health search to others as 65%. 50% of women are more reasonable than men to say their latest search was at least in part of others as 62% compared to 50% of men. The characteristic most likely to define a health seeker on favour of others the good health of the user is 59%healthy. Internet users look on favour of others compared to 32% of those in poor health. And 62% of 30-49 year-old did research for someone .Compared to last time 38% of Internet users aged 65 and over.

Some e-patients use their new found awareness of health resources to advice for loved ones, taking them to doctor's appointment and connecting them to other people with the same diagnosis. One person said that, "Being informed makes it easier for me to be a support to the family and friends in a time of need." Another person said that online health information helped to relax some of the husband's affair about cancer surgery. "It helped to be ready for a new way of life," "He has also get benefits from others who have interest with the same task." A second person from our report is the small group of Internet users who are living with a disability, handicap t, or chronic disease. Fifteen per cent of Americans says in our survey that a disability, handicap t, or chronic disease keeps them from concentrating wholly in work, school, housework, or other activities. The tendency of disability and chronic illness increases with age (5%) of 18 – 29 year-old live with a disability or chronic illness and the rate increases with age up to (28%) of Americans over 65 years old. Americans those who have a disability have the lowest levels of Internet access in the country. In a survey, in the month of spring 2002, we found that 38% of Americans with incompetency go online, compared to 58% of Americans. They have less

confidence than other non-users to believe that they will never use the Internet and less likely than others to live physically and politically close to the Internet. Americans those who have a disability are also less likely to have friends or family who are in online.

A third group that appeared from our reports are those who take care of others living in their household.11% of Americans live with others who is chronically ill or disabled and 70% of that group is a primary or secondary caregivers. Six million of home caregivers go online and, when looking for health information, tend to focus on duties like treatments, procedures, or drugs. Stimulated home caregivers are more similar than the general Internet user population they have searched for information about a particular medical treatment or procedure (62% vs. 47%) and for information about prescription or about the drugs (55% vs. 34%). They are also more similar than the general Internet users who have researched mental health information (37% vs. 21%), physical treatments (35% vs. 18%) and Medicare or Medicaid (21% vs. 9%). Caregivers in general, avoid online health researchers – they are just as similar as non-caregivers to have searched for all the other topics we discussed. Care giving duties are more similar for older generations than to Americans under 30. Just 15% of 18-29 year-old who live with a care recipient say they are primarily responsible for the helpless loved one, compared to about 60% of Americans above the age of 30 who are living with a care recipient.

4. INTERACTIVE HEALTH SERVICES

To make more formatted decisions towards better health, health seekers are getting increasingly savvy with their information needs. Specifically, each health seeker has a very specific need and knows what they expect when they look into the Internet. This leads to diverse, sophisticated and complex motivations and needs of online health seeking. To gain insights into health seeker needs, we randomly collected 5000 Q and A pairs from Health Tap, which cover a wide range of topics, including cancer, endocrine and pregnancy. We carefully went over all these Q and A pairs and observed that the health seeker needs can be enlarge into three main categories. Specific motivations and question examples are also provided to enhance the understanding of this categorization. It can be seen that the three categories do not mutually overlap and cover all the possible cases. This is because the health seeker with respect to a concerned health problem can only be in one state out of the three at one time healthy status suffering from diagnosed disease or undiagnosed disease.

A user study to investigate the health seeker needs. Three volunteers were invited to manually classify each of the 5000 Q and A pairs into one of the three predefined categories. It is worth notifying that each volunteer was well trained with the definitions of category types as well as corresponding examples. We performed a voting method to establish the final classification of each Q and A pair. For cases where each class equally receiving one vote, a discussion was carried out among the volunteers to obtain the final decision. According to our statistics, the distributions of Q and A pairs over the three categories are 79, 6 and 15 per cent, respectively. Even though the third category is not the majority, it greatly increases the bottlenecks of the automatic health system as we have analysed before. Automatically categorizing this community generated health data is somewhat difficult because of the unknown language and gap. Regarding the disprove languages; negated identifiers are frequently used by medical practitioners to indicate that patients do not have given conditions. Some traditional approaches do not discuss between the positive and negative contexts of medical concepts in medical records, which may prevent the learning/retrieval performance from being effective. Take the following two short medical records as an example. Intuitively, their contexts are totally different, while a learning or search system may inaccurately consider such medical records to be equivalent

In the health communities, users with different backgrounds do not necessarily share the same terminology. Sometimes, the same medical subjects may be commonly expressed with various medical concepts. For example, “birth control” and “family planning” are commonly used by individuals to refer to the same medical rules “contraception”. The traditional context representation such as n-gram is unable to capture the variants of medical concepts and may lead to an explosion of feature dimension. To tackle these problems, we employed the Meta Map tool detect medical attributes that are noun phrases in health domain and then normalize them to standardized terminologies in the SNOMED CT Met thesaurus. In our previous work, we have detailed this procedure. The semantic types of these terminologies span from symptom, treatment, meditation, body parts, to demographics. In this paper, we utilize these normalized medical attributes to represent the community generated health data. We represent the Q and A pairs with these terminologies and study the health seeker needs via Q and A pair classification.

5. DECISION SUPPORT SYSTEM FOR DISEASE ANALYSIS

The sparse deep learning schema is enhanced to fetch discriminate features from health data values. Medical terminology based gerontology is used for inference estimation process. Feature analysis is carried out with the conceptual relationship based weight values. Question and Answer data values are evaluated with symptom priority levels. The disease inference estimation scheme is constructed to analyse the question and answers in online health services. Medical domain based gerontology is adapted to identify the disease inferences. Feature selection and categorization operations are integrated with the system. The system is partitioned into six major modules; they are Question and Answer Sessions, Tag Analysis, Investigation with gerontology, Deep Learning Process, Discriminatory Feature Selection and Inference Identification Process.

The Question and Answers (Q A) session module is designed to perform the data p reprocess. Tags are identified and categorized under the tag analysis. Question and Answer data values are analysed with gerontology. Features and signatures are identified under deep learning process. Discriminatory feature identification is used to discover

features for decision making process. Disease discovery is carried out under the inference identification process.

The Question and Answer (Q A) data sets are collected from online health services. The Q and A data values are transferred into the database. Questions, answers and tags are extracted from the data sets. The data sets are labelled with category information. The tags and associated disease information are identified in the tag analysis. Overlapped tag details are also updated with category information. Keyword feature identification is performed in the tag analysis. Features and associated tags labels are updated into the database. Medical gerontology is constructed with disease categories and term elements. Term feature relationship is also represented in the gerontology. Q and A data values are analysed with gerontology elements. Keyword features are assigned with conceptual relationship weight values. Pseudo labelled data and doctor labelled data are analysed in the learning process. Signatures are identified from the raw features under the global learning process. Input layers and hidden layers are updated with features and signatures. The layers are used in the inference identification process.

Unstructured community generated data values are analysed to fetch the discriminatory features. Features are ranked with the conceptual relationship based weight values. Symptom priority levels are also used to identify the discriminatory features. Overlapped features are also verified in the feature selection process. Automated disease inference identification is carried out for the community based health services. Sparse deep learning based inference identification process is applied without concept relations. Concept hierarchy is used in the Gerontology Supported Space Deep Learning method. Healthy status, diagnosed disease and undiagnosed disease labels are produced in the inference identification process.

6. PERFORMANCE ANALYSIS

The Question and Answer (QA) based decision support system is constructed to identify the discriminations based on the Patient and Doctor communication. The Sparse Deep Learning (SPL) mechanism is applied to estimate the disease status for the patients. The Ontology is used to maintain the concept and term relationships. The Ontology supported Sparse Deep Learning (OSDL) scheme is built with medical gerontology and Sparse Deep Learning scheme. Symptom priority levels and discriminatory feature identification mechanisms are integrated with the Ontology supported Sparse Deep Learning (OSDL) scheme. The system performances are compared with the Sparse Deep Learning (SDL) scheme and the Ontology supported Sparse Deep Learning (OSDL) schemes.

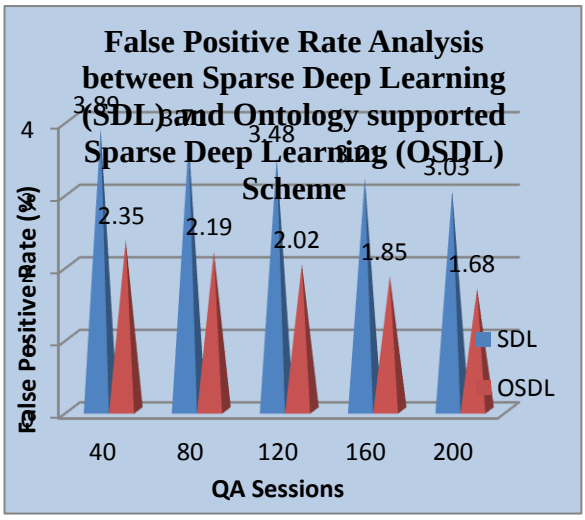

Figure No: 6.1. False Positive Rate Analysis between Sparse Deep Learning (SDL) and Ontology supported Sparse Deep Learning (OSDL) Scheme

The disease diagnosis decision support system is tested with two performance measures. They are false positive rate and false negative rate. The false positive rate and false negative rate measures are employed to estimate the decision making accuracy level of the system.

The false positive rate analysis is estimated with the positive discriminatory results and the falsely assigned positive results. Figure 6.1. Shows the False Positive Rate analysis between the Sparse Deep Learning (SDL) scheme and Ontology supported Sparse Deep Learning (OSDL) schemes. The analysis result shows that the Ontology supported Sparse Deep Learning (OSDL) scheme reduces the False Positive Rate 30% than the Sparse Deep Learning (SDL) scheme.

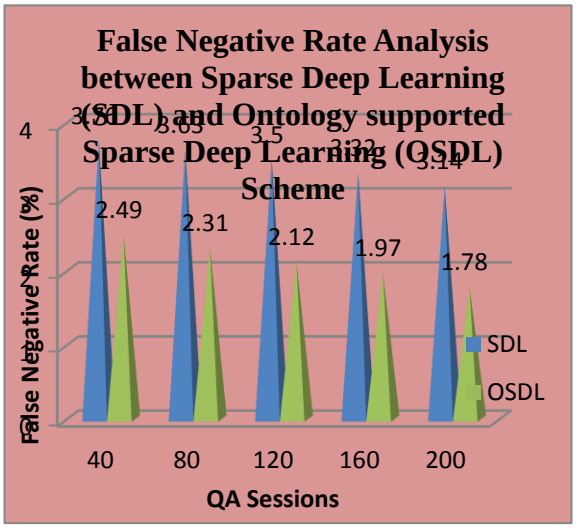

Figure No: 6.2. False Negative Rate Analysis between Sparse Deep Learning (SDL) and Ontology supported Sparse Deep Learning (OSDL) Scheme

The false negative rate analysis is estimated with the negative discriminatory results and the falsely assigned

negative results. Figure 6.2. Shows the False Negative Rate analysis between the Sparse Deep Learning (SDL) scheme and Ontology supported Sparse Deep Learning (OSDL) schemes. The analysis result shows that the Ontology supported Sparse Deep Learning (OSDL) scheme reduces the False Negative Rate 25% than the Sparse Deep Learning (SDL) scheme.

7. CONCLUSION AND FUTURE WORK

Online health services are deployed to provide remote medical assistance. Automatic disease inference estimation is carried out using Question and Answer (Q A) based diagnosis details. Sparse deep learning scheme is improved with gerontology support. Discriminate feature identification mechanism is used to upgrade the inference estimation process. Efficient discriminatory feature identification model is adapted in the community based health services. The system provides Gerontology support for Question and Answer (Q A) based communication process. The Question and Answer (Q A) based system performs decision making with feature priority values. The system improves the accuracy in inference estimation process. Audio/Video based QA session analysis and multi session analysis methods can be integrated with the decision support system in the future development process.

REFERENCES

[1] LiqiangNie, JialieShen and Tat-Seng Chua, "Bridging the Vocabulary Gap between Health Seekers and Healthcare Knowledge", IEEE Transactions On Knowledge And Data Engineering, Vol. 27, No. 2, February 2015

[2] Y. Chen, Z. Chenqing and K.-Y.Su, "A joint model to identify and align bilingual named entities," Comput.Linguistics, 2013.

[3] P. Bansal, P. MacConnachie and O. James, "Bridging the research–practice gap," Acad. Manag.Perspectives, vol. 26, pp. 73–91, 2012.

[4] N. Chu, Y. Choi, J. Wei and A. Cheok, "Games bridging cultural communications," in Proc. IEEE Global Conf. Consumer Electron., 2012.

[5] Z. Fu, G. Lu, K. M. Ting and D. Zhang, "A survey of audio-based music classification and annotation," IEEE Trans. Multimedia, vol. 13, no. 2, pp. 303-319, Apr. 2011.

[6] E. J. M. Laura 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, 2011.

[7] Tom Chao Zhou, Michael R. Lyu and Irwin King "A classification-based approach to question routing in community question answering", ACM, 2012.

[8] B. Li and I. King. Routing questions to appropriate answerers in community question answering services. In Proceeding of the ACM 19th conference on Information and Knowledge Management, pages 1585–1588, 2010.

[9] X. Si, E. Chang, Z. Gyongyi and M. Sun. Confucius and its intelligent disciples: Integrating social with search. Volume 3, 2010.

[10] Y. Yan, G. Fung, J. G. Dy and R. Rosales, "Medical coding classification by leveraging inter-code relationships," in Proc. ACM SIGKDD Int. Conf. Knowl. Discov.Data Mining, 2012.