IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2320-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 5, May 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.512

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 10, Issue 5, May 2021||

DOI:10.15680/IJIRSET.2021.1005077

Question Answering System Using Natural Language Processing

Akash Thombare, Sonal Waikar, Manish Mishra

Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune, India

ABSTRACT: The important data on the health related questions and answers on websites can be helpful to patients and doctors. On this site noisy questions/answers demand medical knowledge and answer pairs are extracted and not related or even wrong information gets filtered. Looked with the huge size of data created on sites and responding to questions of medicine every day, it is impossible to play out this activity with a supervised method due to the extreme budget annotations. So, we projected a system of Medical Knowledge Extraction (MKE) that can spontaneously generate high quality entity extracted from noisy couples of questions and answers and also evaluate the experience of the doctors who suggest answers. The MKE system is based on a framework to discover the truth, which together calculate the reliability of the answers and the physicians experience from the data without supervision. Furthermore, address three one-of-a-kind difficulties for the undertaking of extricating restorative information, i.e the portrayal of confusion, different associated certainties and long tail marvel in the information. Both quantitative and contextual investigations exhibit that the proposed MKE framework can give helpful medicinal information and exact therapeutic experience. Likewise, we exhibit an application in real world dataset, ask a specialist, who can consequently offer recommendations to patients about their inquiries.

KEYWORDS: trustworthiness degree(TD), Medical Knowledge Extraction(MKE), Crowd sourced Question Answering (CQA).Question answering (QA), Question (Q) and Answer (A)

I. INTRODUCTION

In modern world, the youth prefer to search their health related problem on various online sources, or ask the doctors through the online platform. The modern system brings new challenges as well asnew opportunities in medical field to patients, doctors and pharmacy sector. Relating to our old type of face-to-faceservices, the new type of health care questionanswering (QA) system gives us a platform to communicate openly with number of people at a time, which increases data tremendously. In this system any patients ask question and many doctors give an answer on that question. Then the system checks the truth discovery of that answer using entity extraction. In the previous, valuable information of crowd-sourced QA websites how to use in real time is the big question. Therefore, we extract knowledge from such information and use it to develop different types of health services. From that knowledge we can construct the robot doctor for answering the new health related questions.

The main challenge of extracting knowledge from the medical crowd-sourced QA system is that the QA sets are not trusted. Hence,we create supervised taking in strategies that remove information from boisterous publicly supported information on therapeutic QA sites. The important segment in the therapeutic learning extraction errand is to discover dependable suggestion with no supervision. Some current truth revelation strategies expect that the appropriate responses from high-ability clients are dependable, and the clients who give solid answers ought to have high aptitude. In this manner these techniques can iteratively appraise the source unwavering quality (i.e., client aptitude) and surmise solid answers from client contributed information. Along these lines we create unsupervised taking in strategies that separate learning from uproarious publicly supported information on therapeutic QA sites. The important segment in the therapeutic information extraction assignment is to discover dependable answers and suggestion with no supervision. Truth disclosure strategies are produced for organized information (e.g., social database), however for publicly supported QA sites, every one of the sources of info are unstructured information (i.e.,writings). We address this test by changing unstructured writings to substance-based portrayal.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 10, Issue 5, May 2021||

DOI:10.15680/IJIRSET.2021.1005077

II. LITERATURE SURVEY

Using the novel technique to code the heath care records, it jointly utilizes local mining and global learning accesses, which are tightly linked mutually reinforced. For extracting medical and health care concepts from heath care records,mining tries to code the individual health care record and then maps them to terminologies. Sometimes, local mining access, suffer from data and information loss, which are caused due to the absence of key health care concepts and the presence of inappropriate medical and health care concepts.[1]

First of all, the user knows the health seekers and their questions, and then for those further analysis, select those who are thinking about possible diseases. And then propose novel deep learning plans to define possible diseases of health seekers questions. The plan to be suggested consists of two main parts. Initially of all, the world's leading natural ingredients in raw features eat signatures. In the secondly,raw features and its signatures are the input nodes. Learning interrelationships between these two levels through pre-training with pseudo-labeled data.[2]

A new problem arises which is known as Veracity. Which means conformity to truth, and in it, it's studied that how truth facts are discovered from a big inconsistent data on many parts produced by distinct systems. They generate a basic structure for veracity problems and construct analgorithm "truthfinder", which uses the system and the relation in to their data and information, that means the information is provided if the website is trusted, and be true if a task of the data and information is produced by number of strong systems. A repeating method is used to understand the reliability of the system and the accuracy of each other's information.[3]

A novel access that deals with dependence among data sources in to the truth discovery. All at once, if two data sources produced a large number of probable values and many of these values are rarely produced by other sources, it is very likely that one replica from the other. They used Bayesian analysis to decide dependence between sources and structure an algorithm that repetitively find dependence and recognize truth from conflicting data.[4]

In that they find the different algorithm for the prior knowledge as framework. A frameworkexpresses all common "common-sense" logic important cases known to the person as first-order logic construct into tractable linear program. As solution see that, this perspective is also good with big problems that reduce the error and allow a majority to make the user know the truth in relation to truth. [5]

To combat the dispute number of origins of heterogeneous types of data. They detect the difficulty in use of the optimization structure, where truth and source credentials are defined as two sets of unknown currencies. Purpose attaches to total weighted deviation of truth and multi-source observation where each origin is loaded with its credibility. Distinct frameworks can be incorporated into this framework to identify the features of different types of data and the effective calculation technique have been construct.[6]

Truth discovery approaches have different assumptions about input data, relationship in to the object's sources and, find out truths, etc. functions in some types of domains also have their rare characteristics that should be consider into account. These challenges the uses fordiverse truth discovery method.[7]

In that they focus on binary measurements. Optimally, in that cases greatest likelihood evaluation, is attain by solving the expectation greatest query or problem that replay the best answer regarding the truths of each measurement. [8]

In that the information about Truthfulness, Internet information in to the two field where they concluded information are kind of clean and information grade is essential to user'slife.Perform two information group state of art data fusion technique that ambitionat resolving conflicts generating the firstly truth, secondly evaluate their strengths and drawback, give suggestion research orientation.[9]

A automated technique innovated the effective of people developed medical and health related statements and thetrustworthiness by performingphilologist cues and distant guidance and supervision from specialist origin and sources. They produced a feasibility guidance model that learns people statement credibility, trustworthiness and language objectivity. This technique to the part of essence-sing rare or unknown side-effects of health care tablet and being one of the most issue where large-scale non-specialist information has potential to complement specialist health related knowledge.[10]

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 10, Issue 5, May 2021||

DOI:10.15680/IJIRSET.2021.1005077

III. PROPOSED METHODOLOGY

In the proposed system, a MKE (Medical Knowledge Extraction) system, conduct the MKE and specialist aptitude estimation with no supervision. In medical crowd-sourced QA systems, patients ask questions with some intentions. For example, patients want to ask the disease related to their symptoms and side effects of tablets. Specialist giveanswers to these questions those play the essential part in the QA systems. Many doctors and specialist may provide separate remedies as per their knowledge to same type of questions. The truth discovery method also proposed in proposed system and this method use to automatically generate doctors and specialist expertise level and arrange weighted aggregation depending on the generated doctor and specialist expertise level. For performing the truth discovery method, startlingly take values from various means and convert texts to the entity-based structure. In truth discovery method, all output terms of containing knowledge of medical field, diagnosis, questions and trustworthiness and calculate doctor's level of expertise. In real-time many application are developed using above outputs.

A. Architecture

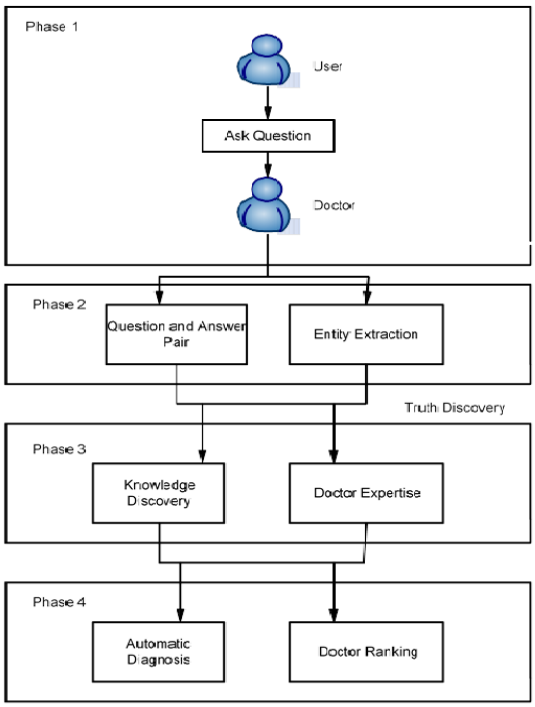

Fig. 1. System Architecture

B. Algorithms

Medical Knowledge Extraction (MKE):

The MKE system extracts the information from noisy data that is Question/answering couple and encapsulate into medical knowledge. After that apply the truth discovery method for finding the truth answer. For that design the 3 module that is Patient, Doctor, and Admin. The patients are asking questions and see the answers. Doctors give the answers to the patient's questions. And the admin works on truth discovery method, checks the truth answer and sends to user. And admin also can see all patients and the doctor's data.

Input: a group of health care problems i.e. Q and their compatible solutions i.e. A; an external entity dictionary with entity class, and real-value vector presentations of entities.

Output: discovered and find knowledge three levels,first question, second diagnosis and third trustworthiness degree, and doctors and specialist expertise level wd.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 10, Issue 5, May 2021||

DOI:10.15680/IJIRSET.2021.1005077

Steps:
1: preliminary procedure: fragment content to words;
2: Extracting Entity:Take one kind of existence (condition)from Q text and another from A type existence (its disorder);
3: input three level creation: form three level <entities from Q text, entity from A text, doctor ID> as input to truth discovery;
4: initiate doctors and specialist expertise level evenly.
5: For calculation is done by using the real value vector representing the entities, where the similar function is used as trusted value according to database;
6: Evaluate doctor expertise wd according to solution and answer
7: return discovered knowledge three level first question, second diagnosis and third trustworthiness degree and the evaluated doctor expertise wd.

C. Mathematical Model

Finding the trustworthiness of answers anddoctor and specialist expertise level as given [3]:
Evaluation of the trustworthiness degree of a desirable answer xq for the q-th question:

$T(x_q) = \sum_{d \epsilon D} w_d.1(x_q, x_q^d)$.....(1)

Updating in doctor and its expertise level:

$w_d = -log(1 - \frac{\sum_x \epsilon V_d T(x)}{|V_d|}$.....(2)

Now,calculate trustworthiness degree and its possible answer is foundas below:

$T(x_q) = \sum_{d \epsilon D} w_d.1(x_q, x_q^d) + \sum_{x_q^1 \neq x_q} Sim(V_{x_q}, V_{x_q^1}).T(x_q^1)$.....(3)

Now add a term for pseudo count as Cpseudo for each and every source and then the expertise score is estimated:

$w_d = -log(1 - \frac{\sum_x \epsilon V_d T(x)}{|V_d| + C_{pseudo}}$.....(4)

In above equation, if a doctor gives less number of answers, then Cpseudo will dominate the term,Vd+ *Cpseudo*, therefore doctor expertise score becomes less. Similarly, if doctor gives more numberof answers, then Vd term gets effected,and thus the score will be nearby to its actual score.

IV. CONCLUSION

In that, medical crowd-sourced QA systems provide valuable information but noisy health related information. Medical Knowledge Extraction system is proposed, to abstract valuable quality of health care inputs i.e knowledge from theQA topic. The MKE systemcan abstract inputs from all QA pairs, expertise knowledge fromtrustworthy sources without any overseeing. The main tasks in extracting knowledge of medical field are to identify and tackle in MKE system. Identifying the similar questions and merge it accordingly and removing noisy inputs we have used entity-basedrepresentation. For connection between the responses, we have used similarity function concept to the model. And to conduct the long phenomenon in the sources, a pseudo count is added which is used to estimate the reasonable medical experience of each doctor. In real-time application, MKE system has great auspicious to advantages more appliance such as artificial doctors or ROBOT.As number of users will increase the system will become more useful.

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 Transactions on Knowledge and Data Engineering, 2015.
2. L. Nie, M. Wang, L. Zhang, S. Yan, B. Zhang, and T.-S. Chua, Disease inference from health-related questions via sparse deep learning, IEEE, 2014.
3. Y.Li, C.Liu, Nan Du, Wei Fan, Qi Li, Jing Gao, C.Zhang, and H.Wu, Extracting Knowledge from crowd-sourced Question Answering website, IEEE,2017
4. X. L. Dong, L. Berti-Equille, and D. Srivastava, Integrating conflicting data: The role of source dependence, The Proceedings of the VLDB Endowment (PVLDB), 2009.
5. J. Pasternack and D. Roth, Knowing what to believe (when you already know something), International Conference on Computational Linguistics (COLING10), 2010
6. Q. Li, Y. Li, J. Gao, B. Zhao, W. Fan, and J. Han, Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation, International Conference on Management of Data (SIGMOD14), 2014

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 10, Issue 5, May 2021||

DOI:10.15680/IJIRSET.2021.1005077

7. Y. Li, J. Gao, C. Meng, Q. Li, L. Su, B. Zhao, W. Fan, and J. Han, A survey on truth discovery, arXiv preprint, 2015.
8. D.Wang, L. Kaplan, H. Le, and T. Abdelzaher, On truth discovery in social sensing: A maximum likelihoodestimation approach, inInternational Conference on Information Processing in Sensor Networks (IPSN12), 2012
9. X. Li, X. L. Dong, K. B. Lyons, W. Meng, and D. Srivastava, Truth finding on the deep web: Is the problem solved? The Proceedings of the VLDB Endowment (PVLDB),2012.
10. S. Mukherjee, G. Weikum, and C. Danescu-Niculescu-Mizil, People on drugs: credibility of user statements in health communities,ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD14), 2014

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.512

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details