

Privacy preserved Recommender system for Breast Cancer Patients

^[1] Sreevidya Nelli, ^[2] Neela Kiranmai

^[1] Asst. Prof, Dept. of IT, SNIST, ^[2] Asst. Prof, Dept. of CSE, SNIST

Abstract: Some of the problems cannot be expressed by some people, especially women suffering from breast cancer. So they cannot express their views publicly and because of this, problem keeps on going and increases the tumour inside there by deteriorating the health totally. The main concern of the patients diagnosed with a medical condition is, which physician is to be chosen. Along with this, they also look for a recommendation system which can protect private information in a secure manner which is a major concern of an interactive user who can share their private information in the internet. So we would like to provide some recommendations to such kind of women who can express their problems privately in a blog, which can be maintained securely. After identifying the women who are suffering or diagnosed with breast cancer, we do some recommendations like:

- Recommendations are provided online, doctor facility to elaborate their problem in detail
- Recommendations based on their symptoms and other relevant predictions of breast cancer.
- Giving the details of other patients (name will be hiding and only revealing the patient symptoms and how they undergo treatment) who are also suffering with the same problem.

Index Terms—big data analytics, recommendation system, security, sentiment.

1. INTRODUCTION

Diagnosis and Prognosis are the two major challenging aspects which are to be addressed in treating breast cancer. The survival of breast Cancer patients depends upon the diagnosis of Cancer at the early stages (either in Stage I or Stage II). If the cancer diagnosed in Stage III or later stages, the chances of survival of the patient will become more critical. Prognosis will reveal the survival pattern for different attributes i.e., for specific drug, before and after the treatment.

Big data technologies are widely used in biomedical and health-care informatics research. Scientists related to biomedical field [7] are facing new challenges day-by-day to maintain, store and analyze the data. Analytics needs novel techniques to analyze the data to provide security to the users. Data security is a major concern while collecting the data and retrieving in an interactive environment. Medical information retrieval in an interactive environment will play an active role while the patients are looking for an active recommendation system. We would like to collect the data online and do analysis using Sentiment Analysis, through which we can classify which women are suffering and their past history for predicting their ancestor's medical history.

To seek medical attention, people usually rely on friends and relatives to choose a physician because of trustworthiness. But, when the patient has no friends or

moves to a new location, they will try to seek help from internet where ratings can be obtained for a physician for a particular medical condition. The main concern of the patients diagnosed with a medical condition is, which physician [1] is to be chosen. Along with this, they also look for a recommendation system which can protect private information in a secure manner. This is the major concern for an interactive user who can share their private information in the internet.

When the patients look for recommendation systems online, they mainly go with the ratings while they are not trustworthy and inaccurate. So to collect the opinions or views from such women who are suffering from such kind of problems, generally single patient records will generate a large amount of data if we manage and analyze such big data, we may solve many problems in identifying the patterns which will lead to diagnose and prognosis of the cancer. This will help the doctors to take proper decisions. In this work, classifying stages of patients, clustering of patients based on their location [2] which will analyse the open Cancer patient's data. We will give possible recommendations based on their stages, location, online doctor facility and online medicine suggestions.

II. SCIENTIFIC BACKGROUND:

Big data applications [6] present new opportunities to discover new knowledge and create novel methods to improve the quality of health care. The application of big

data in health care is a fast-growing field, with many new discoveries and methodologies.

Need for Big Data Analytics in Healthcare [3]

To improve the quality of healthcare by considering the following:

Providing patient centric services: To provide faster relief to the patients by providing evidence based medicine detecting diseases at the earlier stages based on the clinical data available, minimizing drug doses to avoid side effect and providing efficient medicine based on genetic makeup's. This helps in reducing readmission rates thereby reducing cost for the patients.

Detecting spreading diseases earlier: Predicting the viral diseases earlier before spreading based on the live analysis. This can be identified by analyzing the social logs of the patients suffering from a disease in a particular geo-location. This helps the healthcare professionals to advise the victims by taking necessary preventive measures.

Big data in health care [4] is associated with huge volume of patient-specific data. Patients can join social networks through which they can exchange information and provide support to each other. Big data is serving consumers more reliably and giving timely information about quality of health care. Primary objective of Big Data Analytics in health care is to improve the efficiency and quality of health care service. Patients can be diagnosed and treated successfully by analysing diseases at earlier stages.

Sentiment can be described as state of the mind, it conveys passion and one's feeling over a subject or topic. Sentiment analysis however, uncovers the attitude conveyed by the writer or speaker as it relates to the topic. Sentiment Analysis [5] in health care is exploding enormously which can assist the health care organizations in improving patient experience. By combining patient data with their satisfaction value and sentiments, one can get the detailed insights in health care domain.

To determine the sentiment of a document and to discover the opinions, Mukras in Representation and learning schemes for sentiment analysis noted that the fundamental method of classifying a document is by counting the number of positive and negative words in the document and then making a decision based on the resulting majority. This analysis gives patients perception towards the disease in health care domain.

Sentiment Analysis applied to patient data is a systematic study of online information and patient's data on satisfaction surveys. Information that is collected is broken into bags of words. These are further analysed and classified according to the stages of cancer, meaning and intensity like how and whether they are positive, negative and neutral based on their interest to express the past medical diagnosis. Based on the opinions obtained from patient's online data, we give recommendations to patients based on the study conducted on variety of techniques related to Big Data Analytics in health informatics.

III. RELATED WORK

The paper deeply explores how sentiment analysis can be applied to opinions extracted from social networks and to demonstrate how sentiments can impact on their treatment for proper diagnosis and we will predict the behaviour of patients and propose efficient recommender system to facility timely treatment. There exists various types of recommendation systems which provides patients to collect dependable doctor recommendations for a medical condition. They are protected in two ways by maintaining privacy of

- 1) Patients giving ratings to the system.
- 2) Patients seeking information.

The recommendation system allows the patients to give their ratings to the extent of satisfaction they get from a physician for a particular medical condition. To protect the physician's ratings being affected by an untruthful user, we ensure that the person is minimally influenced and the physician also can not alter the ratings.

There has been considerable research in privacy preserving recommendation systems. Initially, privacy was achieved in recommendation system by transmitting use information to a trusted third party which performs necessary calculations with another trusted agents.

Recommendation system must be capable of withstanding against certain attacks like Shilling Attacks, Bad Mouting, and Boosting.

Shilling Attack is one in which misbehaving users may try to affect the rating of a particular physician. Detecting such attacks is a future area of research. Detection algorithms mainly focus on system abuse rather than abuse prevention.

Bad mouthing is another attack where the patient try to decrease the rating of a physician.

Boosting occurs when the patients collaborate to increase the rating of a particular physician.

To avoid the need for a trusted -third party we have a common way based on Homomorphic Encryption and Data Perturbation. In Homomorphic encryption the users encrypt their data before sending to a trusted-third party so that the hidden plaintext is not visible to any third party. This is how the user’s privacy is preserved since no other party will not be able to decrypt the encrypted text. Another approach is data perturbation in which the users manipulate their user-data by adding noise to it before using in any computation.

Thus the user’s original ratings are secured. Another approach related to privacy used in recommendation system is Differential Privacy in which the function is not primarily affected by any minor changes to the database. Information is collected online and then submitted to Sentiment Analysis techniques to analyse the data as either positive, negative or neutral. After collecting the data, it is pre-processed and their semantics are analysed and the overall sentiment is calculated. Then the desired Recommender System will be ready for predicting the online patient data. To protect sensitive information related to medical records there exists two frameworks

1. Secure Processing Architecture (SPA).
2. Anonymous Contributions Architecture (ACA).

In SPA[6] patients offer encrypted ratings and the calculations of all recommendations is done on secured data through secure multiparty computation in which multiple computational servers gather data from patients and compute recommendations as shown in fig. 1. In ACA as the name indicates, the patients use anonymous routing for submitting their contributions to the entity that receives all patient ratings and publishes information about them. Today so many anonymizer systems like Tor anonymity network [7] and other anonymizer services and proxies such as Anonymizer Inc. and Shadow surf [8] are available.

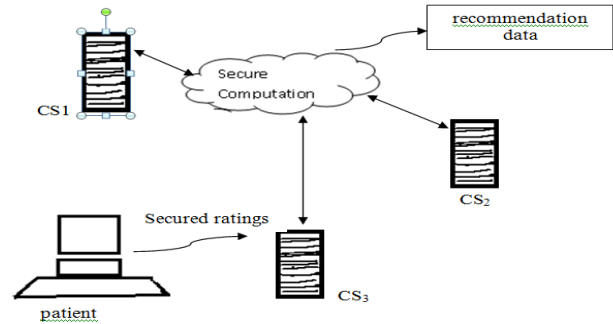


Fig 1. Submission of User Ratings and recommendation computation in SPA architecture

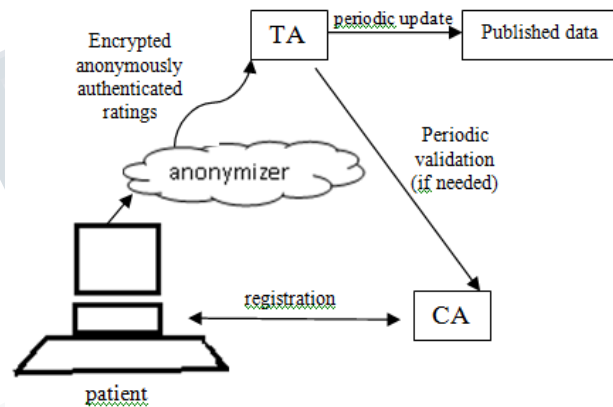


Fig 2. Submission of User Ratings and recommendation computation in ACA architecture

As described in fig. 2 A Patient who wants to contribute their ratings interacts with CA during registration only and the person will be given anonymous credentials. The remaining interaction occurs with the TA and the person who is querying must access published data only. Even though there is communication between CA and TA or not to authorize the patient’s credentials, the querier will be able to decide a ranked list of physicians for single and multiple medical conditions.

There is a second solution for implementing ACA based on electronic cash in which anonymity can be maintained and even the double spending can be found. Because our requirement is the user anonymity who is giving ratings and duplications on same physician can be found. So we use these two techniques, for the purpose of maintaining anonymity to the user.

Although SPA and ACA meet all the requirements for calculating recommendations for a better medical system, both of them suffer from certain flaws. They are summarized below:

Demerits of SPA

1. Estimating the recommendations includes substantial computational load on the servers while using multiparty computation techniques. New recommendations must be produced rarely when the load depends on existing physicians in the system as well as number of health conditions it supports.
2. Sometimes users might have queries related to specific medical conditions, such types of queries cannot be executed due to high computational costs on the servers. In addition to that, such type of queries may leak some information related to conditions included in the query.
3. System cannot guarantee privacy with distrustful parties to perform the computations.

Demerits of ACA

1. Explaining the concept of anonymous authentication to average user becomes difficult, if they would like to give contributions, they must be willing to register.
2. When the system is used by less number of users, privacy weakens. These kind of vulnerabilities must be known to the users of the system.

IV. CONCLUSION

The outcome is to provide recommendations to the patients who are hesitant to express their personnel details online for a better health care, in this process we are reducing both the costs incurred for treatment, time spent to seek medical attention and successfully getting medicinal facility to overcome the disease. So finally by providing these recommendations, we can make sure that they are seeking medical help without hesitation. we conclude that there is a need for an efficient Recommendation system which can overcome the drawbacks that are present in the existing approaches mentioned above, that can guarantee user privacy in all the possible ways and avoid duplications while giving ratings to a particular physician. We propose a system in which Security can be further enhanced so that the users can rely on the system what they are looking for and ensure that privacy for each patient is maintained throughout and they can go to doctor easily. The proposed

system works efficiently and accurately predicting the online patient's medical conditions.

REFERENCES

- [1] Reliable Medical Recommendation Systems with Patient Privacy T. Ryan Hoens^{1,2}, Marina Blanton¹, Nitesh V. Chawla, Copyright 2010 ACM 978-1-4503-0030-8/10/11.
- [2] BERJANI, B. AND STRUFE, T.2011. A Recommendation system for spots in location-based online social networks. In Proceedings of the EuroSys workshop on Social Network Systems
- [3] Promises and Challenges of Big Data Computing in Health Sciences Tao Huang, Liang Lan, Xuexian Fang, PengAn, Junxia Min, Fudi Wang in Big Data Research Journal (2015), <http://dx.doi.org/10.1016/j.bdr.2015.02.002>.
- [4] A Survey Of Big Data Analytics in Healthcare and Government J.Archenaa and E.A.Mary Anita, 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15), Elsevier.
- [5] Twitter Sentiment Analysis in Health Care by using Hadoop and R, Vijay Shankar Gupta, ShruthiKohli, 978-9-3805-4421-2/16 ©2016 IEEE.
- [6] BAUDRON, O.FOUQUE, P.-A, POINTCHEVAL, D.STERN,J.,AND POUPARD, G.2001.Practical multi-candidate election scheme. In Proceedings of the ACM Symposium on Principles of Distributed Computing (PODC).274-283
- [7]BLANTON, M.2008. Online subscriptions with anonymous access. In Proceedings of the ACM symposium on Information, Computer and Communications Security (ASIACCS).217-227
- [8] Big Data Application in Biomedical Research and Health Care: A Literature Review