

A Framework of Clinical Recommender System with Genomic Information

Qian Zhang, Guangquan Zhang, Jie Lu

Center for Artificial Intelligence, Faculty of Information Technology

University of Technology Sydney, NSW, Australia

E-mail: {qian.zhang-1, guangquan.zhang, jie.lu}@uts.edu.au

Hua Lin

23 Strands, 105/26 Pirrama Rd, Sydney, NSW 2009, Australia

E-mail: hua.lin@23strands.com

Clinicians make decisions that affect life and death, quality of life, every single day. It is important to support clinicians by discovering medical knowledge from the accumulated electronic health records (EHRs). The integration of genomic information and EHRs are long recognized by the medical community as the inherent feature of the disease. The demand for developing a clinical recommender system that is able to deal with both genomic and phenotypic data is urgent. This paper proposes a framework of clinical recommender system with genomic information, which is used in the clinical process and connects the four types of users: clinicians, patients, clinical labs, researchers. With models and methods in artificial intelligence (AI), five functions are designed in this framework: diagnosis prediction, disease risk prediction, test prediction, and event prediction. The proposed framework will help clinicians to make decisions on the next step in clinical care action for patients.

Keywords: Recommender system; Clinical recommender system, Genomic information, Precision medicine.

1. Introduction

Due to the large number of patients clinicians are meeting every day, efficiency and accuracy of decision making are quite crucial in clinical medicine¹. Because of the fast variation of diseases, numerous new drugs are developed to deal with different situations occurred to patients. It is crucial to keep up to date with plenty and complex disease and drug information for clinicians. For example, contraindications of many drugs may bring chronic harm to unsuitable patients or what's worse may cause lethal effect. To help clinicians reduce mistakes and improve efficiency in making decisions of the next step clinical care action for patients, there is a requirement of developing clinical recommender systems.

The accumulation of electronic health record (EHR) provides the opportunity of building clinical recommender system. EHRs are patient-centered records containing demographics, clinician orders, test results, diagnosis, procedures etc. in the clinical process². EHRs provide a multi-view of clinician behavior and patient information in clinical process and mainly collected through an EHR system. The start of clinical support is by adding smart searching strategies into the system. Inforbuttons is a system that allow clinicians to search electronic resource collections through queries based on patient data³. Further, a decision support system called lung cancer assistant is developed to assist clinicians in treatment selection by case-based reasoning and probabilistic models⁴.

Disease inheritance has been long recognized by medical community, and the genetics is the cause and contributor to the complexity of clinical medicine⁵. If the cause of symptoms is not clearly understood, it is not possible to provide the proper treatment to patients. For example, patients are vulnerable of developing life-threatening myelosuppression if they harbor genetic variants that prevent them from metabolizing the immunosuppressive drug azathioprine normally⁶. The challenge of developing clinical recommender system is how to properly deal with the genetic information of patients. Personalized medicine (now more widely known as precision medicine) is a term that raised to contain the sequencing of human genome in clinical medicine. The link of genomic data and EHRs help to understand deeper on the cause and therapies of the diseases⁷. Thus, developing clinical recommender system with genetic information is to use genomic approach to tailor the clinical process to prevent the disease and improve the healthcare experience of patients.

To solve the above problems, this study proposes a framework of clinical recommender system with genetic information. Firstly, how the clinical recommender system framework will work in the clinical process is presented from four user perspectives. Then, how genetics helps in the framework on predictions of diagnosis, disease risks, test, prescription and other events are presented.

The rest of the paper is arranged as follows. It introduces the related works in Section 2. Section 3 explains the clinical process and the framework of clinical recommender system. In Section 4, the paper discusses future work and concludes this study.

2. Related works

This section demonstrate related work about clinical decision support systems and describe how this paper fits into the scope.

Clinical decision support “refers broadly to providing clinicians and/or patients with clinical knowledge and patient-related information, intelligently filtered, or presented at appropriate times, to enhance patient care.”⁸ The clinical decision support can be grouped into three categories: passive, asynchronous and active⁹. Passive decision support refers to non-mandatory resources during the clinical process. The EHRs and related system can be seen as an example of passive decision support. The easy access to these data to clinicians during the consultation and care of the patient provides support for clinicians. Asynchronous decision support involves system monitoring, data assembling, report generating and email notifications. From the genetic view, some testing such as germline genetic test is only needed once in a lifetime. The asynchronous decision support system is to monitor and inform clinicians to avoid duplicate testing. The active clinical decision support is to provide information such as probability of diagnosis, medicines to prescribe and tests to order during the clinical process. One example of active clinical decision support is to raise alert when the clinician has given a risky decision on prescribing a drug that is allergy for a patient. The study of this paper is on the basis of the three types of decision support, and provide personalized recommendation for clinicians.

3. A Framework of Clinical Recommender System

This section describes the clinical process overview between the patients, clinicians, clinical lab and AI researchers. Followed by the overview of the clinical recommender system and the function of each component of the framework.

3.1. *Clinical process: overview and perspectives*

This part gives an overview of how the clinical process goes when a patient comes and how the clinical recommender system is involved in this process in Fig. 1.

When a patient comes to a consultation, the clinician will give the next action for the patient, and use the symptoms, test results, family records of the patient to help him make the decision. The clinician needs to determine the cause of the symptom and patient’s prognosis. To do this, next action such as clinical test or genetic test may be needed and all data need to be cooperatively gathered to make the final decision. The clinical recommender system is to help clinicians quickly, efficiently and accurately predict the most appropriate action for a patient, such as order a test, diagnosis, prescription or hospitalization. The aim of the clinical recommender system is to simplify the complexity of the clinical process, lower the risk of clinician’s decision and managing the information from several sources.

Obviously, the process involves four kinds of people that are related to the clinical recommender system. We will give four perspective from different users:

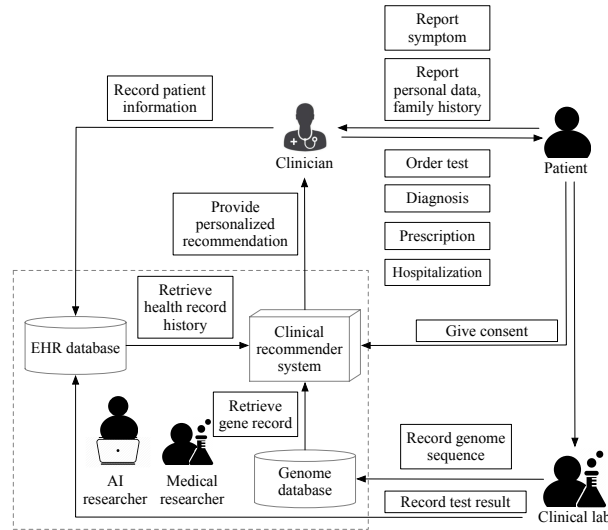


Fig. 1. Clinical process of patient

(1) Patient perspective: The role of patients is becoming more important since they are now able to access their genetic data directly from some testing providers. There are more and more concerns on the health privacy and the data sharing. So the participate of patient and their consent to access their data are crucial for building the clinical recommender system. It is essential to contain a portal for patients to make it easier for them to effectively and efficiently consent to share their data in supporting research and development of medical health. The patient portal is needed in the system to allow patients to visualize their health condition, access their full medical data and control their privacy.

(2) Clinician perspective: The role of clinician is the core part in the whole process to give patients suggestion of next step action. Clinicians need to combine all the information and interpret the symptom and disease for the patient. The interpretation are needed from both phenotypic and genomic side: (1) Phenotypic side: Clinicians determine which clinical test might be needed, how the test results can reveal the cause and severity of the symptoms and what interventions are needed for the patient. (2) Genomic side: Clinicians determine which variants in the genome are related to the disease, how the variants affect the symptoms and how the variants affect the treatment or clinical care for the patients. The interpretation and decision of clinicians contain professional judgement therefore human error and subject difference will affect the quality and efficiency of the

clinical process. The recommender system is truly needed to lower the risk and reduce the number of errors by providing personalized recommendation for clinicians.

(3) Clinical lab perspective: The clinical lab provides the results mainly on the genomic test, but also the clinical test such as blood test. Clinical labs provide all the variants in the genome and their possible relationships with the disease. There are many existing variants that are already known with specific impact, but more variants are to be discovered and interpreted. The data structure of a combination between the clinical test data in EHR database and genomic data in genome dataset is needed for the development of the clinical recommender system. An example of existing public dataset containing a total of 158,668 variants submitted by 369 organizations * is ClinVar from National Center for Biotechnology Information (NCBI).

(4) The AI researcher perspective: Artificial intelligence (AI) is developing fast in the era of big data. With both genomic and phenotypic data, there are many methods and models to be developed to fit various requirements and provide different support to clinicians, or even to medical researchers. The collaboration of AI and medicine is needed in establishing gold standard for the relationships between phenotypic features such as Prostate cancer and genomic features such as single nucleotide polymorphism. Knowledge base should be established as basis of building the clinical recommender system, which definitely involves the participation of medical experts. Opportunities also lie in the discovery of new medical knowledge especially with the genomic information is exploited.

With incremental data accumulated through each patient consultation and interaction, the clinical recommender system is incrementally improved and will benefit the clinician, patients, lab and medical researchers. The ultimate goal is to give personalized recommendation to support decision making of clinicians and improve the healthcare of the patients.

3.2. Framework of the clinical recommender system

This part proposes an overview of the framework of clinical recommender system with genomic information as shown in Fig. 2. Since each disease has their specific symptoms or features, the clinical recommender systems should be built separately on each disease.

The basis of the clinical recommender system is the preprocessing and analysis of the data from both genomic data and phenotypic data respectively:

* The data is recorded in 11 September 2015.

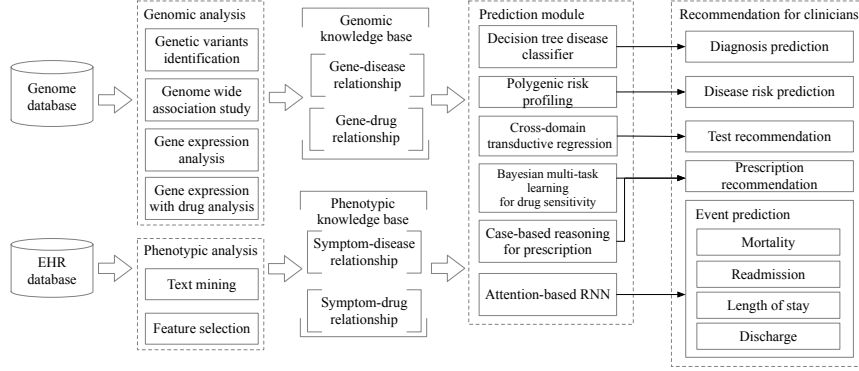


Fig. 2. Framework of clinical recommender system

(1) Genomic data analysis: As shown in Fig. 2, sequencing data from individuals are stored in genome dataset. The genomic data analysis contain two levels. The first is the analysis on genetic variations. Genetic variations contain two types: single nucleotide polymorphisms (SNPs) and Structural variation (SV). Identification of variants connected to specific diseases is the basis of genome wide association studies (GWAS). GWAS are methods for identifying associations between genetic regions and traits (including diseases). A typical GWAS study is to use statistical methods to find the relationship between genetic variants and diseases. Except for the level of the gene copy numbers, the gene expression level is not only related with disease, but also the sensitivity of drugs. GWAS together with gene expression analysis the main component to establish the genomic knowledge base.

(2) Phenotypic data analysis: The main dataset for the phenotypic data is the EHR of patients which contains patient demographics, family history, clinician orders, diagnoses, procedures, prescriptions, lab test values, hospitalization flowsheet and free-text medical notes. Feature selection is needed to extract the symptoms and other related features from structured data. Also text mining is needed for the unstructured data such as free-text medical notes. With features and disease labels, case-level knowledge bases are established.

There are five functions in the framework as listed in the right part. Each function may be connected with one or more prediction models or methods. Here, we briefly introduce one representative method that may achieve the function, and emphasis how genetic information plays a part in each function.

(1) Diagnosis prediction: To make diagnosis, features about symptoms or tests are extracted from structured test results or unstructured electrocardiogram. Decision tree is able to select features and classify the patients to the appropriate type of the disease. Genetic variation associated with high likelihood of disease

manifestation. Genetic variants may reveal the cause of disease and provide evidence for confirmation of diagnosis.

(2) Disease risk prediction: This function aims at early disease detection, prevention and intervention for asymptomatic individuals. Family history helps to show the genetic inheritance of some diseases. Also, genetic variants affect and lead to the disease not by single loci but multiple loci. Polygenic score is a number from regression analysis on variation in multiple genetic loci with associated weights. The utility of polygenic risk profiling is used to predict the probabilistic susceptibility of a disease, and patients are classified to high, intermediate and low risk to take different actions. Except gene variants, gene expression is also highly related with the disease risk.

(3) Test prediction: The test prediction is mainly on the genetic sequencing test. Since the multiple genetic loci are functioning together in many diseases, cross-domain regression and recommendation are needed to infer the relationships between multiple generic loci. The recommendation results may suggest exome sequencing or whole genome sequencing and analysis. Sequencing of parents or other family members may also need.

(4) Prescription prediction: The symptoms, diagnosis and lab test results determine the prescription of a clinician to patients. The prescription prediction can be solved by case-based reasoning with phenotypic data. Also genetic variants and gene expression affect the patient's respond to drug therapy. Multitask learning is effective to deal with complex molecular feature space with multiple drugs.

(5) Event prediction: The EHR can be processed and converted to events such as test, therapy or discharge ordered in a timeline. The deep learning method especially the attention-based recurrent neural network (RNN) use the time series events to make prediction at specific time points. The genetic variants may affect mortality and discharge prediction.

From the above framework, the understanding of genetic variants and gene expression can be considered in conjunction with EHR data and support the decision making of clinicians on the next step action in various aspects. The participation of genetic data in the clinical recommender system is crucial to improve the healthcare for the patients.

4. Conclusion and future work

This paper proposes a framework for clinical recommender system to support clinicians in decision making for providing patients with better healthcare. How the clinical recommender system is working in the clinical process is explained

from four user perspectives. In the framework, the genomic data and phenotypic data are processed with GWAS or AI techniques and the genomic data affect five functions in prediction of diagnosis, disease risks, test, prescription and other events.

Future studies will focus on implementation of each part in the framework and how the datasets of genomic and phenotypic sides can integrate with one another. Furthermore, domain knowledge of experts in medicine and bioinformatics is needed to evaluate the system before it comes to practical use.

Acknowledgments

We acknowledge the 23Strand company for background information.

References

1. Lu, J., Wu, D., Mao, M., Wang, W. and Zhang, G., 2015. Recommender system application developments: a survey. *Decision Support Systems*, 74, pp.12-32.
2. Zhang, Q., Zhang, G., Lu, J. and Wu, D., 2015, November. A framework of hybrid recommender system for personalized clinical prescription. In *2015 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)* (pp. 189-195). IEEE.
3. Cimino, J.J., 2008, November. Infobuttons: anticipatory passive decision support. In *AMIA Annual Symposium Proceedings* (pp. 1203-1204).
4. Sesen, M.B., Peake, M.D., Banares-Alcantara, R., Tse, D., Kadir, T., Stanley, R., Gleeson, F. and Brady, M., 2014. Lung cancer assistant: a hybrid clinical decision support application for lung cancer care. *Journal of The Royal Society Interface*, 11(98), p.20140534.
5. Aronson, S.J. and Rehm, H.L., 2015. Building the foundation for genomics in precision medicine. *Nature*, 526(7573), pp.336-342.
6. Relling, M.V., Gardner, E.E., Sandborn, W.J., Schmiegelow, K., Pui, C.H., Yee, S.W., Stein, C.M., Carrillo, M., Evans, W.E. and Klein, T.E., 2011. Clinical Pharmacogenetics Implementation Consortium guidelines for thiopurine methyltransferase genotype and thiopurine dosing. *Clinical Pharmacology & Therapeutics*, 89(3), pp.387-391.
7. Abul-Husn, N.S. and Kenny, E.E., 2019. Personalized medicine and the power of electronic health records. *Cell*, 177(1), pp.58-69.
8. Hoffman, M.A. and Williams, M.S., 2011. Electronic medical records and personalized medicine. *Human genetics*, 130(1), pp.33-39.
9. Osheroff, J.A., Teich, J.M., Middleton, B., Steen, E.B., Wright, A. and Detmer, D.E., 2007. A roadmap for national action on clinical decision support. *Journal of the American medical informatics association*, 14(2), pp.141-145.