Int J Basic Sci Med. 2016;1(1):13-17 Original Article

http://ijbsm.zbmu.ac.ir/ doi 10.15171/ijbsm.2016.04

Dynamic Recommendation: Disease Prediction and Prevention Using Recommender System

Mahdi Nasiri[1*], Behrouz Minaei[1], Amir Kiani[2]

[1]Computer Engineering Department, Iran University of Science and Technology (IUST), Tehran, Iran
[2]Department of Mathematics and Computer Science, Amir Kabir University of Technology, Tehran, Iran

***Correspondence to**
Mahdi Nasiri;
Email: nasiri_m@comp.iust.ac.ir

Published online 29 June 2016

Abstract

Background: In today's world, chronic diseases are predominant health problems and cause heavy burden on society; therefore early diagnosis and even prediction of the disease is a way to reduce this burden. In this project, we tried to use recommender system to predict which other diseases a chronic patient is susceptible for.

Methods: In this study, through a dynamic recommender system, we evaluated patients' treatment destiny during the time.

Results: It was shown that our method increased accuracy and reduced error compared with other recommendation methods in disease prediction.

Conclusion: Compared to current usual methods, in our method we used previous patients' characteristics as one of the factorization variables to predict destiny of future patients. Furthermore, using this method, we can predict which complication or disease the patient would suffer from first in future. Therefore, we can manage policies toward disease burden reduction by implementing prevention programs.

Keywords: Recommender system, Disease prediction, Collaborative filtering, Data mining, Treatment

Please cite this article as follows: Nasiri M, Minaei B, Amir Kiani A. Dynamic recommendation: disease prediction and prevention using recommender system. Int J Basic Sci Med. 2016;1(1):13-17. doi:10.15171/ijbms.2016.04.

Introduction

Cost of diseases and strategies to decrease it is one of the most important problems. In some of illnesses, the process of treatment is too long and the annual cost of treatment is pretty high. Besides, chronic diseases are long-term diseases that restrict patients' physical activity. They are difficult to cure, need long-term treatment, and process improvement is difficult. In some cases, there is no specific cure for the disease. Therefore, the use of modern techniques such as data mining can be effective to reduce the time and cost of this type of diseases. One of the usages of data mining is in recommender systems.[1-4] The system gives recommendation for the treatment of disease based on three elements including patients, treatment, and time.[4] Resnick et al and Shardanand & Maes[5,6] provided a substantial review on the current system deficiencies and potential advantages of health system. This research suggested that data mining is the main field of prospective health care.

It is undeniable that genomics research is rapidly expanding, though the general applicability is limited.[7] Hence data[8] showed that data and existing technology can provide immediate progress toward the prospective medical research. Risk factors can be combined and the impact would be high risk susceptibility of disease,[9] and the incidence of both diseases would raise.[10] There are many different methods of computing for medical prediction. A known system, Apache 3, which is a scoring system for predicting mortality associated with the disease and predicting future impacts, uses a combination of acute physiological measurements, age, and chronic health conditions.[11,12] Some recommender system algorithms use linear algebra algorithms.[13,14] The recommender system uses fuzzy rule based algorithm[15,16] and clinical factorization for increasing the performance. In addition, evaluation and a recommendation engine (CARE) uses ICD-9-CM codes based on a patients' medical history.[17-19] CARE also refines participatory methods to predict the highest risk of each patient based on his/her medical history and sim-

Copyright © 2016 The Author(s); Published by Zabol University of Medical Sciences. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ilarity. However, using this algorithm has the problems such as data distribution and cold start point.

Recommender system tries to offer personalized recommendation from a large number of possible options out of modeling features list. These systems filter the information on the web and provide information tailored to the interests of their users. In general, two approaches are used in the recommender systems: content-based approach and collaborative filtering (CF)-based approach.

Content-based approach proposes the user, looking for items, the content (properties) that would have more similarity to the content of items in the past and given preference by the target user. The CF-approach uses other users' ratings to predict the target users' unknown ratings. One of the advantages of this approach relative to content-based approach is that in this approach we do not need specific proposals and knowledge and the approach can be used to offer various items such as movies, photos, music, etc. The implementation of this approach is also simpler than that of content-based approach. In addition, content-based approach needs to identify and exploit content features that are often ambiguous and maybe these characteristics are not useful to predict users' interests or to make a distinction between useful items. Moreover, the CF-approach is appropriate to cover undetected patterns whose detection in content-based approach is difficult or impossible.

Most techniques used in a collaborative algorithm (CA)-approach to create the model are as follows: Bayesian classification, neural networks, fuzzy systems, and matrix factorization. Matrix factorization techniques are the most successful and famous model-based techniques. One advantage of this technique is that despite the sparseness matrix ratings, recommender systems provide for high prediction accuracy. Matrix factorization techniques try to describe user features from hidden factors of users and items. These factors automatically are derived from the feedback of users (known as points). In these techniques, users and items are mapped to a hidden operating environment. We used the spaces as latent factors to predict the unknown users' ratings.

In this study, we aimed to offer appropriate treatment for chronic patients who are treated during time. The reason that why we used CF methods is that the treatment for each patient is different from others; and we could find a patient who had similar treatment and could suggest specific recommendations for other patients.

Methods

CF systems produce their proposals on the basis of information obtained from similar users. As a result, contents of items are not considered. In this research, by user we mean patient, and by item we intend the susceptibility to chronic diseases. For prognostic prediction of a patient with a particular disease, we can use the estimation of individuals having similar disease. In this way, the basic assumption is complications of a certain disease for similar patients that are repeatable in future. Categorizing similar patients is important, as one single patient could suffer from many overlapping diseases during his/her lifetime and many different diseases could be presented with similar signs and symptoms. These problems can influence proper collaborative filtering.

In this paper, in a dynamic recommender system we evaluated patients' treatment destiny during the time. Time is an important aspect in the treatment of chronic diseases. In fact, we faced a three-dimensional data in which the third dimension was time. Therefore data was stored in a three-dimensional array which is called tensor. The question is how to work with a three-dimensional object in order to create a recommender system. This modeling allowed us to have different proposals at different times for different individuals. Modeling is presented based on tensor factorization that is extended from matrix factorization for three dimensions or more. Innovative articles use tensor decomposition model in recommender system. In this article, we tried to solve the three dimensional tensor of patient-disease-time. In a similar two-dimensional mode, we can use the following factorization based on HoSVD algorithms. HoSVD is a common tensor decomposition algorithm. If we have only two dimensions, namely disease and patient, we can have this formula:

$$\hat{r}_{u.i} = b_u + b_i + p_u.q_i^T \quad (1)$$

Some methods add some features like means of ratings as in formula (2). b_u and b_i are means of users rating and items rating.

$$\hat{r}_{u.i} = b_u + b_i + p_u.q_i^T \quad (2)$$

We tried to decompose a tensor R with three dimensions of patient-disease-time to three latent matrices of patient (U), disease (I), and time (T). As noted, it is calculated based on the HoSVD algorithm. Patient latent matrix includes hidden features of patients that are results of tensor decomposition. We can formulate our problem as:

$$R = (U, I, T).S \quad (3)$$

Since the calculation results from the patient, disease, and time matrices due to the plenty of missing data factors, it cannot be equal to the basic R tensor, and the results are approximate and therefore a method should be provided to minimize the approximation and to be able to calculate the real amount.

$$R \approx (U, I, T).S \quad (4)$$

In above formula, U indicates patient vector matrix, I indicates disease vector matrix, and T indicates time vector matrix. In fact, the S is the tensor called core. The matrices of patient, disease, and time dimensions are orthogonal. The number of decomposed matrices including patient, disease, and time is respectively equal to the number of patient, disease, and time of basic tensor. And the column

scan varies. The best value of matrix columns (factors) that is defined in the result as K, is obtained by calculating different values. In fact, the value of k as the final choice is to minimize the error in estimating the approximate amount of R. Here, patient and disease are supposed to be independent factors and time as an independent risk factor shows the feature of this approach. The proposal for a disease (l) and a patient (u) and time (t) by using the HoSVD factorization is placed in the following formula:

$$r_{uit} = \sum_{m}\sum_{n}\sum_{l}(u_{um} i_{in} t_{tl}) S_{uit} \quad (5)$$

In the formula (5), if U_m, is mth row of the matrix U, I_n, nth row of the matrix I, and T_l, lst row of T, then we can estimate the value of the tensor cloning according to the following formula:

$$r_{uit} = (U^T{}_u, I^T{}_i, T^T{}_t).S \quad (6)$$

However, the results of factorization of HoSVD may increase the suggestion for the information that we did not have. Like two-dimensional mode, we did not want to rebuild elements that got very large by this method, so we had to control the analytical columns metrics in a way that these elements did not get too large. Then in the case of tensor factorization, we could replace the HoSVD factorization with the minimizing ordered issue and get the proper answer.

$$Min \sum_{(u,i,t)\in k} (r_{uit} - \overline{r_{uit}})^2 + \lambda(\| U_u \|^2 + \| I_i \|^2 + \| T_t \|^2) \quad (7)$$

The second part of the above formula was to avoid sudden jumps and value of λ determined between 0 and 1. It should be noted that the number of elements were rated to *k* using the above model. The tensor empty value which was the proposed treatment of patients is identified. The point that should be noted was that in this method, the value of tensor is a binary content and the content of the new obtained matrix were values between 0 and 1, which using a threshold would convert to binary content.

Results

Hospital recorded data are used as reliable and efficient references for disease identification. Scattering data in a data set from the equation (8) can be obtained. In this regard, |R|, |U|, and |I| by means of rates, number of users, and the number of items are desired data collected.

$$\text{Sparsity} = 1 - \frac{|R|}{|I| \times |U|} \quad (8)$$

Distribution of Medicare data, according to the amount of votes, the users and the items in it and calculated based on the equation (8) was equal to 0.99. Since this number was closer to 1, the distribution of this data set was very large, showing appropriate sampling. It was also very convenient for testing the recommender algorithms which were going to overcome the vote matrix scattering.

Assessing the recommender system means measuring the system's accuracy in predicting future additive disease of a certain patient with a certain current disease. Accuracy and recall are two criteria which were used to evaluate the prediction performance. For diseases that were specified, 80% as training data and 20% as testing data were evaluated.

This was repeated for five times. The average error of about five times was considered as total error. We compared our method with other algorithms like ICARE. Results showed that our algorithm had better results on this data. One of the algorithms covered 100% of data. All cells of matrix of patients-diseases were set between 0 and 1. We named these numbers as risk of any disease on any patient. Physicians need to predict some diseases for any patient and for this reason we used top n diseases for any patient. We used top 20 diseases with high risk. We also used time that meant sequence of visits and helped us to predict future high risk diseases. Table 1 shows that for 20 top higher risk diseases, our algorithm had more coverage in comparison with ICARE and other models that did not use time dimension. Accuracy of disease prediction for patients was also higher than others.

Another point for using the time dimension was higher accuracy and less error for last time. Figure 1 shows that time dimension reduced error of disease prediction.

Discussion

Chronic diseases are predominant health problems. These diseases are associated with economic, social, cultural, emotional, and many other aspects of human life. Each single disease either mild or sever have its certain burden on society and its prevention, diagnosis, and treatment can reduce this burden. Burden of disease is measured by disability adjusted life years and it

Table 1. Comparing Top 20 Higher Risks

	Using Time Dimension	Without Time Dimension	ICARE
Coverage	63.2%	56.2%	41.2%
Accuracy	76.21	68.52	49.272

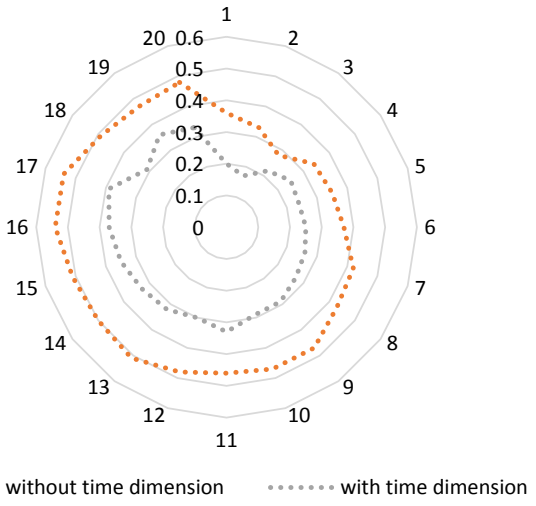

Figure 1. RMSE Error for Proposed Method With Different Factors.

influences gross domestic production in countries.[21-24] If either communicable or non communicable disease leads to chronicity, it would cause heavier burden on society. Body will be under a long time stress in chronic disease; this would prepare proper situation for new diseases appearance. Therefore in patients with chronic disease, there are trends for catching new additive diseases. There are many studies on how to control disease chronicity and its complications. Bodenheimer et al[25] and lorig et al[26] showed that self examination is a good method for management of chronic disease and related complications. There are plenty of epidemiological studies performed to demonstrate disease habits, vulnerable groups, and related complications. Moreover, some studies tried to predict patients' destiny, and disease future and stages. Clayton et al[27] calculated family tendency to some diseases using epidemiological analytical models. Intelligent artificial networks are also used to predict treatment outcomes in many diseases.[28-33] In this project, we used recommender system to predict which other diseases a chronic patient is susceptible for. We used three dimensional analytical algorithms using patient, disease, and time as its factors. In comparison with current usual methods, in this method we used previous patients' characteristics as one of the factorization variables to predict destiny of future patients. Furthermore using this method, we can predict which complication or disease the patient would suffer from first in future. Therefore, we can manage policies toward disease burden reduction by implementing prevention programs.

Conclusion

Early detection of chronic disease leads to prompt treatment, reduces treatment cost, and increases the chance of complete remission. In this method, using recommendation algorithms based on tensor factorization, the opportunity of predicting future additive diseases on top of current disease of a certain patient is provided. The results of the proposed method were evaluated on the data. In this survey, 80% of data were selected as training data and 20% were selected as test data and results had high accuracy in diagnosis for patients before or at early chronic disease. The proposed method is an algorithm, based on tensor factorization, which anticipates the disease for any patient. By forecast, it is meant the time period that specifies in what stage the illness would be, not the exact time of the disease. Matrix and tensor factorization algorithms among recommender system algorithms have better behavior on sparsity challenge which is a big challenge. Solving tensor decomposition based on optimization methods reduces the error of rating prediction on collaborative filtering methods. In collaborative filtering algorithms, user and item similarity are used for prediction. These algorithms, in comparison to other data mining algorithms can more often solve data sparseness and that is its great challenge. The general approach to solve is to predict the rates based on final patient similarity to its neighbors. Tensor factorization based on the optimization algorithm is used to solve and help to reduce algorithm forecast error and improve user-centric collaborative refinement. Development and evaluation of the proposed system show that using the recommender system is a strong and appropriate approach for prognosis and recommended treatment. It is also recommended that for future work, tensor factorization procedures to be used to reduce the sparseness of data and to increase the accuracy of diagnosis. Use of tensor factorization method can be considered as the possible risk of disease rather than a definitive decision. Other characteristics such as geographic location and family history can also be examined for more accuracy.

Ethical Approval

Not applicable.

Competing Interests

Authors declare that they have no competing interests.

References

1. Esmaeili L, Nasiri M, Minaei-Bidgoli B. Personalizing Group Recommendation to Social Network Users. In: Gong Z, Luo X, Chen J, Lei J, Lee F. Wang Web Information Systems and Mining. Berlin Heidelberg: Springer; 2011: 124-133.
2. Loscalzo J. Association studies in an era of too much information - clinical factorization of new biomarker andgenetic data. Circulation 2007;17:1866-1870. doi:10.1161/circulationaha.107.741611.
3. Paterek A. Improving regularized singular value decomposition for collaborative filtering. KDDCup; August 2007.
4. Nasiri M, Rezghi M, Minaei B. New algorithm for recommender system based on tensor decomposition. Journal of Operational Research in Its Applications. 2014;14:57-64.
5. Resnick P, Iancovou N, Sushak M, et al. Grouplens: an open architecture for collaborative filtering of netnews. In: Proceedings of the ACM Conference on Computer Supported Cooperative; 1994. p. 175-186.
6. Shardanand U, Maes P. Social information filtering: algorithms for automating "word of mouth". In: Proceedings of the computer human interaction; 1995. pp. 210-217.
7. Si L, Jin R. Flexible mixture model for collaborative filtering. Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003); Washington DC; 2003.
8. Starfield B, Lemke KW, Bernhardt T, et al. Comorbidity: implications for the importance of primary care in case management. Ann Fam Med. 2003;1:8-14.
9. Akker V, Buntinx M, Metsemakers F. Multimorbidity in general practice: prevalence, incidence, and determinants of co-occurring chronic and recurrent diseases. Journal Clin Epidemiol 1998;51:367-375.
10. Weston AD, Hood L. Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. Journal of Proteome Res 2004;3:179-196. doi:10.1021/pr0499693.
11. Vlachos M, Fusco F, Mavroforakis C, Vassiliadis V. Improving Co-cluster quality with application of product recommendations,. Proceedings of the 23rd ACM

International Conference on Conference on Information and Knowledge Management (CIKM); China; 2014. p. 679-688.
12. Burges C. A tutorial on support vector machines for pattern recognition. Data Min Knowledge Discovery 1998;2:121-167.
13. Sharifi Z, Rezghi M, Nasiri M. Alleviate Sparsity Problem using Hybrid Model based on Spectral Co-Clustering and Tensor Factorization. 5th International eConference on Computer and Knowledge Engineering (ICCKE); Mashhad; 2015. p. 89-91.
14. Christakis NA, Allison PD. Mortality after the hospitalization of a spouse. N Engl J Med. 2006;354:719-730. doi:10.1056/nejmsa050196.
15. Cordn O, Herrera F, de la Montaña J, Sánchez A, Villar P. A prediction system for cardiovas-cularity diseases using genetic fuzzy rule-based systems. In: Proceedings of the 8th Ibero–American Conference on AI; 2002; p. 381-391.
16. Coyle P, Hartung HP. Use of interferon beta in multiple sclerosis: rationale for early treatment andevidence of dose- and frequency-dependent effects on clinical response. Multiple Scler. 2002;8:2-9. doi:10.1177/135245850200800102.
17. Davis D, Chawla NV, Blumm N, Christakis N. Predicting individual disease risk based on medical history. In: Proceedings of the ACM Conference on Information and Knowledge Management; 2008.
18. Edelman D. (2006) A multidimensional integrative medicine intervention to improve cardiovascular risk. J Gen Intern Med. 2006;21:728-734. doi:10.1111/j.1525-1497.2006.00495.x.
19. Shabanpoor M, Mahdavi M. Implementation of a Recommender System on Medical Recognition and Treatment. IJEEEE. 2012;2(4):315-318.
20. Nasiri M, Sharifi Z, Minaei B. Alleviate sparsity problem using hybrid model based on spectral co-clustering and tensor factorization. Computer and Knowledge Engineering (ICCKE), 2015 5th International Conference on,IEEE; 2015. p. 285-289.
21. Yusuf S, Wood D, Ralston J, Reddy KS. The World Heart Federation's vision for worldwide cardiovascular disease prevention. Lancet. 2015;386(9991):399-402. doi:10.1016/S0140-6736(15)60265-3
22. Sargazi A, Sepehri Z, Sagazi A, Jim PN, Kiani Z. Eastern Mediterranean region tuberculosis economic burden in 2014. Antimicrob Resist Infect Control. 2015;4(Suppl 1):P102. doi:10.1186/2047-2994-4-S1-P102.
23. Aryal KK, Mehata S2, Neupane S. The burden and determinants of non communicable diseases risk factors in Nepal: findings from a nationwide STEPS survey. PLoS One. 2015;10(8):e0134834. doi:10.1371/journal.pone.0134834.
24. Sargazi A, Sargazi A, Nadakkavukaran Jim PK, et al. Economic burden of road traffic accidents; report from a single center from south Eastern Iran. Bull Emerg Trauma. 2016;4(1):43-47.
25. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. JAMA. 2002;288(19):2469-2475.
26. Lorig KR1, Sobel DS, Ritter PL, Laurent D, Hobbs M. Effect of a self-management program on patients with chronic disease. Eff Clin Pract. 2001;4(6):256-262.
27. Clayton DG. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika. 1978;65(1):141-151.
28. Sargolzaee Aval F, Behnaz N, Raoufy MR, Alavian SM. Predicting the outcomes of combination therapy in patients with chronic hepatitis C using artificial neural network. Hepat Mon. 2014;14(6):e17028. doi:10.5812/hepatmon.17028.
29. Rathore H. Artificial Neural Network. In: Rathore H, ed. Mapping Biological Systems to Network Systems. Switzerland: Springer; 2016:79-96.
30. Søreide K, Thorsen K, Søreide J. Predicting outcomes in patients with perforated gastroduodenal ulcers: artificial neural network modelling indicates a highly complex disease. Eur J Trauma Emerg Surg. 2015;41(1):91-98. doi:10.1007/s00068-014-0417-4.
31. Abdelwahab A, Sekiya H, Matsuba I Horiuchi Y, Kuroiwa S. Alleviating the sparsity problem of collaborative filtering using an efficient iterative clustered prediction technique. Int J Inf Technol Decis Mak. 2012;11 (1):33-35.doi:10.1142/S0219622012500022.
32. Yang D, Ma Z, Buja A. A Sparse SVD Method for High-Dimensional Data. J Comput Graph Stat. 2014;23(4):923-942. doi: 10.1080/10618600.2013.858632.
33. Nasiri M, Minaei B. Increasing prediction accuracy in collaborative filtering with initialized factor matrices. J Supercomput; 2016;72(6):2157-2169. doi: 10.1007/s11227-016-1717-8.