Recommending best course of treatment based on similarities of prognostic markers

Sudhanshu, Narinder Singh Punn, Sanjay Kumar Sonbhadra, and Sonali Agarwal
Indian Institute of Information Technology Allahabad, India
{ism2016004,pse2017002,rsi2017502,sonali}@iiita.ac.in

Abstract. With the advancement in the technology sector spanning over every field, a huge influx of information is inevitable. Among all the opportunities that the advancements in the technology have brought, one of them is to propose efficient solutions for data retrieval. This means that from an enormous pile of data, the retrieval methods should allow the users to fetch the relevant and recent data over time. In the field of entertainment and e-commerce, recommender systems have been functioning to provide the aforementioned. Employing the same systems in the medical domain could definitely prove to be useful in variety of ways. Following this context, the goal of this paper is to propose collaborative filtering based recommender system in the healthcare sector to recommend remedies based on the symptoms experienced by the patients. Furthermore, a new dataset is developed consisting of remedies concerning various diseases to address the limited availability of the data. The proposed recommender system accepts the prognostic markers of a patient as the input and generates the best remedy course. With several experimental trials, the proposed model achieved promising results in recommending the possible remedy for given prognostic markers.

Keywords: Health recommender system · Prognostic markers · Collaborative filtering · Machine learning

1 Introduction

Recently, recommender systems have become an important part of many different sectors. Major e-commerce platforms employ the use of recommender systems to display the filtered results for every customer. These recommendations keep updating over time with the aim to improve the user’s experience of the platform. The other factors that a recommendation system brings to a platform are: sales boost, enhanced customer engagement, transform shoppers to clients, increase average order value, lower manual work and overhead, and bring more traffic on the e-commerce platform. Recommender systems are also utilized in the following areas: entertainment and media (movie/song/book/news), economy (stocks), banking, telecommunications, etc.

* All authors contributed equally.
The recommender systems are also termed as SaaS (Software as a Service) and hence could be employed in various other fields given the right dataset. Following this, many recommender system based approaches have been utilized in the healthcare domain to aim for better healthcare services. One of the applications is to recommend remedies to a patient that would be most effective in the treatment process. The system proposed in the paper analyzes the prognostic markers for the patient in question by using the collaborative filtering to match the profiles of other patients that had similar prognostic markers. The top remedies will be listed as the output which have proven to be effective on the other patients’ health status when their prognostic markers were on similar levels.

1.1 Motivation and contribution

Recommender systems are being widely used to recommend the relevant items in the context of e-commerce and infotainment. Recommender systems have helped the businesses because they improve the inventory value, user experience etc.

The crux of what is happening behind the scenes is, users are being delivered the relevant content from a huge stack of information. Consequently, when the same approach was applied in the healthcare domain, it gave rise to health recommender systems (HRS). Following this, there was an undeniable boost in automated healthcare and tele-medicine. Health recommender systems (HRS) have been put to use in the following forms:

1. Enterprise Resource Planning(ERP system) 4
2. A Doctor recommendation algorithm 3
3. Web-based RS suggesting online health resources 37

Fig. 1. A typical scenario of a collaborative recommender system.
4. A diet recommendation system to a patient
5. Chronic Disease Diagnosis Prediction and Recommendation System

According to Park et al., the research in the field of health recommender systems has increased immensely but the practical implementations of such systems still requires more research. The major takeaway is that, even with the adequate knowledge in the field of health recommender systems, they are not being put in practical use on a large scale. In this paper, we propose a system in the field of diagnosing and treating diseases in the essence of automating healthcare. The major contributions of the present article are as follows:

1. Developing a recommender system that when given a set of symptoms, will perform a diagnosis and then recommend the next best course of treatment.
2. Creating a data set that contains 'course of treatment' corresponding to their diseases. And these diseases range from mild and acute to chronic states.
3. Provide possible future improvements for recommender systems in the healthcare domain.

### 1.2 Terminologies used

1. **Recommender system (RS)**: recommender systems are prediction algorithms that are aimed at recommending relevant items or information to the users (usually based on their past preferences).
2. **Collaborative recommender systems**: These systems recognize the similarities among the given set of users based on their common reviews and provide updated recommendations based on the similar/contrasting features among the users. Fig. represents a typical scenario where 'User-1' and 'User-2' have liking for 'Item-A', 'Item-B' and 'Item-C'. This makes both the users similar in the context of 'the user preference profile'. Hence when 'User-2' adds a new item labeled 'Item-C', the same is recommended to 'User-1' because it’s highly probable that 'User-1' will also prefer 'Item-C' in the future (as both users have preferred the same things in the past).
3. **Bio markers**: A bio marker, or biological marker is a measurable indicator of some biological state or condition.
4. **Prognostic markers**: Bio markers which can be used to estimate the progression of a medical condition (usually a disease) in an objective manner are called prognostic markers. Prognostic markers are primarily used to divide the patients into categories, aimed at providing precise medicine discovery.
5. **Health recommender systems (HRS)**: recommender systems when applied in the healthcare domain are called health recommender systems.

### 1.3 Organization of the paper

The rest of the paper is divided into several sections. Second section mentions the related works in the field of health recommender systems and the associated
challenges. Third section presents the methodology behind building the proposed system. It also mentions the pre-processing of the dataset, and the inclusion of a newly created data file in the dataset. Fourth section presents the results obtained by the proposed approach. Last section concludes the paper with the outlines of the future scope of recommender systems in the healthcare domain.

2 Related work

The massive growth and advancements in deep learning algorithms across vivid domains such as healthcare, image processing, etc. \cite{7,26,36,40} has resulted in immense applications by developing real-world applications. In the earlier survey of recommender systems by Park et al. \cite{12} it was observed that scholastic investigation on recommender systems have expanded fundamentally over the last ten years, but more insights are needed to develop real-world applications. The research field on health recommender systems is potentially wide, however there are less developments in the practical scenarios. In like manner, the current articles on recommender systems should also be reviewed up to the coming age of health recommender systems. Hors-Fraile et al. \cite{19} also discovered the need of increasing and improving the research in the area of HRS that covers the proposed multidisciplinary taxonomy. This includes the features like integration with electronic health records, incorporation of health promotion theoretical factors and behavior change theories. Kamran and Javed \cite{27} presented a survey of RS in healthcare and also proposed a hybrid recommender system that takes into account the features like hospital quality (measured objectively based on doctors communication, nurses communication, staff behavior, pain control procedures, medicine explanation, guidance during recovery at home, surrounding cleanliness, quietness in patient’s surrounding) and patients similarity. This recommender system suggests the best hospitals for a patient based on the above factors.

Pincay et al. \cite{5} presented a state-of-the-art review providing insights about methods and techniques used in the design and development of HRS(s), focusing on the areas or types of the recommendations these systems provide and the data representations that are employed to build a knowledge base. Sezgin et al. \cite{4} outlined the major approaches of HRS which included current developments in the market, challenges, and opportunities regarding HRS and emerging approaches. Huang et al. \cite{3} proposed an algorithm which improves the performance of the medical appointment procedure. This algorithm creates a ‘doctor performance model’ based on the reception and appointment status. It also creates a ‘patient preference model’ based on the current and historical reservation choices which help in the accurate recommendation. It prevents the situation where a doctor is under-appointed or over-appointed and the patients are not being treated even if doctors are available. Peito \cite{35} proposed a HRS for patient-doctor match-making based on patients’ individual health profiles and consultation history. Another utility HRS was proposed by Kim et al. \cite{1} that personalized diet recommended service through considering the real-time vital sign, family history,
food preference, and intake of users to solve the limitations in the existing diet recommendation services.

Hussein et al. [2] proposed a HRS with the hypothesis that, if a patient’s chronic disease diagnosis and set of medical advice are predicted and recommended with high accuracy, it is expected to reflect the improvement of patients’ health conditions and lifestyle adjustments along with reducing the healthcare services costs. This can be considered as a ‘core health recommendation system’, as it directly focuses on the disease and the preventive side of the healthcare field, whereas other HRS usually help a medical institution function better in other aspects. In another similar work, Kuanr et al. [6] proposed a HRS to help women by providing information on the features responsible for prognosis of cervical cancer in women. Cheung et al. [9] presented another review which outlines that incorporating multiple filtering, i.e. making a hybrid system could potentially add value to traditional tailoring with regard to enhancing the user experience. This study illustrates how recommender systems, especially hybrid programs, may have the potential to bring tailored digital health forward.

Considering the nature of recommender systems, it’s not easy to confine them to some specific sectors. Traditional recommender systems are either collaborative or content-based (broadly speaking). In HRS, which type of recommender system should be used depends on the application. For instance, collaborative filtering might be used in an educational context, whereas content-based filtering would prove to have more impact in creating a doctor recommendation algorithm that takes into account the performance of doctors as well. In collaborative filtering, only the objective information regarding the items are stored, whereas, in content-based filtering, more comprehensive information is stored which gives rise to the following major privacy issue [4]: ‘Combining data from multiple users (probably from different geographical locations) can be seen as an intrusion to the individual private data. It may even uncover some confidential data of healthcare institutions’. This poses a major challenge that violates the delicate topic about privacy which must be confidential in a healthcare system.

Apart from this major healthcare sector confined flaw, in general there are some other basic challenges that a recommender system faces (which are also applicable to HRS):

1. **Data sparsity**: If the recommender system is employed in very few places then the performance of the system may not be very promising. As it will only infer suggestions based on the data samples considered within its limited range, it will not follow into the standard footsteps of a recommender system which usually works on large data samples.

2. **Scalability**: If the users in the system scale to a very large number, say in millions, then the collaborative filtering algorithm will fail. The linear run time algorithm (i.e. $O(n)$ time-complexity) is undesirable in healthcare scenarios because the results should be generated in real-time.

3. **Diversity and the long tail**: The system won’t be able to recommend remedies with a limited amount of historical data. If a remedy is not recommended to a set of patients with similar prognostic markers as the current user, then
that remedy is unlikely to be recommended to the current user, even though that specific remedy could prove more beneficial than the rest.

In the healthcare sector, every peripheral context is as important as working on ailments and their remedies. All five health recommender systems mentioned in Section 1.1, are aimed at providing recommendations in the healthcare context. Among these HRS(s), Hussein et al. [2] approach works on predicting diseases or recommending treatment. The CDD recommender system (as they coined it), acts as a core HRS since it predicts the disease and recommends medical advice. This type of HRS acts as an extra tool, by assisting the physicians and patients in controlling and managing the diseases. They have employed ‘decision tree’ algorithm in ‘random forest’ manner for disease prediction and used a ‘unified collaborative filtering’ approach for advice recommendation. This complete model seems a breakthrough in the HRS sector, however the model is built to predict and diagnose only the ‘chronic diseases’. Following this context, the present research work aims to develop an HRS that:

1. Acts as a core HRS (acting as a tool to help recommend medical remedies).
2. Provides medical remedies to a wide range of diseases, not just the chronic diseases.

3 Methodology

The recommender systems are usually of the following types:

1. Collaborative filtering: It works by locating peer items with a rating history similar to the current item, and then these nearest neighbours are used to generate recommendations.
2. Content-based filtering: It matches the new items for the user with the items previously rated by the same user.
3. Multi-criteria recommender systems: It takes into consideration multiple criterion for suggesting an item to a user.
4. Mobile recommender systems: It aims to provide personalised recommendations while keeping in mind that the mobile data is more complex than the regular data and the protection of privacy needs to be incorporated carefully.
5. Hybrid recommender systems: It works by combining collaborative filtering, content-based filtering and other approaches into one system.
6. Session-based recommender systems: It doesn’t take into account the older history of the user but the usage patterns of current session only.

As mentioned earlier, collaborative filtering and content-based filtering are the most common approaches to build a recommendation system. The proposed model of HRS is built using the collaborative filtering technique.
3.1 Collaborative filtering

Collaborative filtering as the name suggests, employs the use of collaboration. The underlying presumption of this approach is that if an individual A has a similar assessment as an individual B on a context, A is probable to have B’s assessment on a different issue in comparison to that of a randomly picked individual. For instance, a collaborative filtering recommendation system for shopping preferences could make predictions about which outfits and accessories a client would like, given an partial list of that client’s preferences. These preferences may include likes or dislikes, frequency of buying from a particular brand, the average spending amount, etc. It should be noted that, even though these predictions use data gathered from numerous clients, ultimately provides tailored predictions to individual client(s). This contrasts from the simpler methodology of giving a normalized rating for every item of interest.

The analogy to the system proposed in this paper holds as: If a person A was cured by the same treatment as person B given the same set of symptoms, A and B are more likely to be cured by the same treatment for a new set of common symptoms.

3.2 Dataset synthesis

Acquiring the dataset for this system was one of the biggest challenges. There are plenty of datasets publicly available for the healthcare domain but most of them conform only to some specific category of illness (like heart diseases only, nervous system disorder only, etc). This system requires a dataset that contains a list of diseases spread over various domains. The base dataset hence chosen is taken from the profile of P. Larmuseau, Pharmacist at Gaver Apotheek (Harelbeke, Flanders, Belgium). The dataset consists of 8 files in .csv format. The primary data from the dataset contained the information arranged in tuples. Following list shows the labeling (of the tuples) as found in the files of the dataset in the format as (file name: (tuple labels) = (corresponding alias)):

1. sym_t.csv : (syd, symptom) = (Symptom identifier, Symptom name)
2. dia_t.csv : (did, diagnose) = (Disease identifier, Disease name)
3. diffsydiw.csv : (syd, did, wei) = (Symptom identifier, Disease identifier, Weight of the symptom on the disease)
4. prec_t.csv : (did, diagnose, pid) = (Disease identifier, Disease name, treatment course)

A total of 1,167 diseases and 273 symptoms are listed in the dataset. Fig.2 shows the snapshots for the above mentioned files in order.

There was no good data sources available for all diseases (from generic to chronic) and their treatment courses. Hence the data file prec_t.csv (name as in the code repository) is created by exploring several medical websites and resources [8,10,11,13,14,16,18,20,24,28,32,42].

1 https://www.kaggle.com/plarmuseau/sdsort
2 https://github.com/sud0lancer/Diagnosis-Precaution-dataset
Fig. 2. The first three files are from the base dataset. The fourth file with column labels as (Disease Identifier, Disease name, treatment course) is synthesized for the proposed remedy recommendation system.

3.3 Dataset pre-processing

In order to develop the complete dataset for the proposed system, the base dataset was cleaned using the following steps:

1. Dropping the rows if any of the attributes was \( NULL \).
2. Removing unrecognizable delimiters and replacing them with commas(,).

3.4 Building a sparse matrix

The pre-processed data from the data-files is used to create a new matrix which is sparse in nature. For instance, the columns from files: sym_t.csv - (syd, symptom), dia_t.csv - (did, diagnose) and diffsydiw.csv - (syd, did, wei) are transformed into a sparse matrix \( \text{Data}(i,j) \) such that \( \text{Data}[i][j] \geq 0 \) represents the
weight of the $j^{th}$ symptom on $i^{th}$ disease, where higher value represents larger weight of a symptom for a disease and 0 represents that the symptom doesn’t give rise to the corresponding disease. This matrix is considered as the source from where the system will generate the recommendation.

### 3.5 Normalization using BM25 weighting

Normalization was done to calculate the average weight (or importance) of a symptom for a disease in presence of other symptoms. BM25 weighting scheme is used for this purpose. BM25 is considered to be a better version of the TF-IDF ranking functions. Main motive of these functions is to estimate the relevance (score, ranking) of a term in a huge text corpus. These functions employ the frequency and rarity of a term to compute their importance to the corpus. TF-IDF uses the Eq. 1 to compute the relevance score:

$$R_{score}(D,T) = \text{termFrequency}(D,T) \cdot \log \left( \frac{N}{\text{docFrequency}(T)} \right)$$  \hspace{1cm} (1)

Here:

1. $R_{score}(D,T) =$ score of a term $T$ in a document $D$,
2. $\text{termFrequency}(D,T) =$ how many times does the term $T$ occur in document $D$.
3. $\text{docFrequency}(T) =$ in how many documents does the term $T$ occur.
4. $N =$ size of the search index or corpus.

Scores for all such documents in the corpus are added to get the final score (ranking) of a term. In contrast, BM25 adds modifications to compute the ranking score as follows:

1. A document length variable is added, where larger documents are penalised having the same term frequency as those of smaller documents.
2. After the term’s frequency has reached saturation, further occurrences don’t affect its score.

In BM25 \[39\], the ranking score can be computed using Eq. 2

$$R_{score}(D,Q) = \sum_{t \in Q} \frac{f_{t,D} \cdot (k_1 + 1)}{f_{t,D} + k_1 \cdot (1 - b + b \cdot \frac{|D|}{\text{avg}(|D|)})} \cdot \log \frac{N - n_t + 0.5}{n_t + 0.5}$$  \hspace{1cm} (2)

Here:

1. $\sum_{t \in Q} =$ sum the scores of each query term,
2. $\frac{f_{t,D}}{f_{t,D} + k_1} =$ term frequency saturation trick,
3. $\frac{1}{(1 - b + b \cdot \frac{|D|}{\text{avg}(|D|)})} =$ adjust saturation curve based on document length,
4. \( \frac{N-n_t+0.5}{n_t+0.5} \) = probabilistic flavour of IDF.

In this case, the analogy holds as: ‘symptom’ is ‘term’, ‘list of diseases’ is ‘the huge text corpus’. After the normalization, the sparse matrix will now contain updated values. For example, \( \text{Data}[i][j] = 3 \) will have changed to either 2.8736 or 3.1252 (exact values may vary) depending on the effect of the other symptoms on the corresponding disease.

3.6 SVD and cosine similarity

Single value decomposition (SVD) in the context of recommendation systems is used as a collaborative filtering (CF) algorithm. It is used as a tool to factorize the sparse matrix to get better recommendations.

SVD is a matrix factorization technique that splits a matrix into products of two or more matrices such that when these constituent matrices are multiplied back, they will return the original matrix. SVD is a most common method for dimensionality reduction. This implies that when the data is available in many dimensions (i.e. there are a lot of attributes projected in same (or different) directions), it is not easy to infer information due to the curse of dimensionality, however, if the data is reduced to less dimensions, then it would be easier to visualize the data and extract the desired information.

Let \( R \in \mathbb{R}^{m \times n} \) be the original data matrix. Then after applying SVD, \( R \) breaks into the following 3 matrices as shown in Eq. 3:

1. \( U \) is a \( m \times r \) orthogonal left singular matrix,
2. \( V \) is a \( r \times n \) orthogonal right singular matrix,
3. \( S \) is a \( r \times r \) diagonal matrix, such that

\[
R = USV^T
\] (3)

The SVD decreases the dimensions of the original matrix \( R \) from \( m \times n \) to \( m \times r \) and \( r \times n \) by extracting its latent factors. In our case, \( R \in \mathbb{R}^{1145 \times 272} \) is reduced as: \( U \in \mathbb{R}^{1145 \times 50} \), \( S \in \mathbb{R}^{50 \times 50} \) and \( V \in \mathbb{R}^{50 \times 272} \). The matrices \( U \) and \( V \) are used to find the recommendations.

Cosine similarity is the measure of similarity between two vectors. This similarity is calculated by measuring the cosine of the angle between the two vectors which may be projected into multidimensional space. It can be applied to items available in a dataset to compute similarity among themselves. Similarity between two vectors (\( A \) and \( B \)) is calculated by dividing the dot product of the two vectors by the magnitude value as shown in Eq. 4. It is worth noting that the similarity (or CS score) of the given two vectors is directly proportional to the angle between them.

\[
\cos \theta = \frac{A \cdot B}{||A|| \cdot ||B||} = \frac{\sum_{i=1}^{n} A_i \cdot B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \cdot \sqrt{\sum_{i=1}^{n} (B_i)^2}}
\] (4)
In the proposed system, cosine similarity finds the $n$ rows (where each row is represented by a disease) from the decomposed matrices, that have the maximum sum of the symptoms weight (which in turn means the diseases having the maximum matching symptoms). Here $n$ is the number (a manual threshold) of diseases that we wish to generate for the given inputs.

4 Results and discussion

The proposed system was tested in two phases in order to determine the working and the performance of the system respectively. In the first phase, the experiments are conducted keeping in mind the related and unrelated symptoms experienced by the patients that signify the real life scenarios. Sometimes a patient might be experiencing multiple symptoms but most of them hint towards a common disease and in the other cases, those symptoms may be completely unrelated to each other and hence the patient might be having multiple diseases. In the case of related symptoms, the HRS is expected to recommend the remedy for the disease that is most likely to happen because of the given multiple symptoms that are related to each other. And in the case of unrelated symptoms, the HRS must recommend the remedies for all the different possible diseases. Table 1 shows the two cases as mentioned above. ‘Case:1 Unrelated-symptoms’ has an array of ‘symptom_id’ as input. It predicts the probable disease(s) and then recommends the best treatment. Likewise for ‘Case2: Related-symptoms’.

| Case ID | Symptom ID | Symptoms | Most probable disease(s) | Best treatment(s) |
|---------|------------|----------|--------------------------|-------------------|
| Related symptoms | 1 | Upper abdominal pain | 1: Ventral hernia: bulging of the abdominal wall. 2: Diverticulosis: weakening of the large intestine wall. | 1: Eating smaller meals may help prevent bloating and swelling. |
| | 2 | Lower abdominal pain | | |
| Unrelated symptoms | 2 | Lower abdominal pain | 1: Ventral hernia: bulging of the abdominal wall. 2: Vitiligo: loss of skin pigment | 1: Laparoscopic surgery. 2: Photodynamic therapy, Medications: Steroid and Immunosuppressive drug. |
| | 81 | Rash | | |

In both the cases, the system takes in an array of symptom IDs, then it predicts the most probable diseases as a result of the symptoms and then recommends the course of treatment. Fig. 3 shows the output produced by the HRS for both the cases of related and unrelated symptoms.
The second phase consists of further evaluations which incorporates the analysis of the quality of the predictions based on various types of testing. The first level of testing is the ‘sanity testing’. In sanity testing, the dataset is divided into two halves while preserving the uniformity of the symptoms subgroups in the dataset. Both the halves are fed as the input to the system, and the corresponding similarity matrices generated must be as close as possible to the original similarity matrix, i.e., it must show minimal difference in the values (along the diagonal) which will indicate that the similarity matrices are very similar. It implies that there exists least dependency on the data and the type of data that is fed to the system. Euclidean distance is used to find the similarity between the matrices. Henceforth it can be said that the proposed system is un-biased towards data. Matrices $M_1$ and $M_2$ show that the euclidean distance matrix has all diagonal values near 0, implying that the two similarity matrices are similar. $M_1$ is composed of two similarity matrices belonging to full dataset and one of the halves of the dataset, respectively. Similar results hold for the similarity of the other half of the dataset with the full dataset. $M_2$ is composed of two similarity matrices belonging to both the halves of the dataset.

$$M_1 = \begin{bmatrix} 0.3610 & 2.8490 & 3.3920 & 4.8063 & \cdots \\ 2.9624 & 0.3843 & 3.2234 & 3.8602 & \cdots \\ 3.3868 & 3.1321 & 0.3062 & 3.3776 & \cdots \\ 4.8199 & 3.8164 & 3.3867 & 0.2363 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

$$M_2 = \begin{bmatrix} 0.3031 & 2.8507 & 3.3858 & 4.8042 & \cdots \\ 2.8626 & 0.1469 & 3.1463 & 3.8111 & \cdots \\ 3.4013 & 3.1316 & 0.4758 & 3.3871 & \cdots \\ 4.8306 & 3.8184 & 3.3974 & 0.3423 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

The second level of testing is the regression testing, in which we use a subset of the training data to generate the output and match with the training set, which proves that the model is correctly created and has proper similarity matrices. Fig. 3 shows that when given a set of symptoms, system predicts 3 out of 4
diseases which resemble the ground truth (since they had the maximum weight in the training dataset). The remaining 4\textsuperscript{th} prediction is also correct but it had a lower weight in the training dataset. Hence it can be said that the system predicts the most probable disease(s) for a given set of symptoms together, not the most probable disease for each individual symptom(s).

\begin{center}
\begin{tabular}{|l|}
\hline
A = test input = [1, 47, 67, 91] \\
\hline
Most probable disease(s) \\
1115 ['Ventral hernia : bulging of the abdominal wall'] \\
825 ['Vocal cord paralysis : voice box weakness'] \\
760 ['Thyroiditis : thyroid inflammation'] \\
1503 ['Scar'] \\
\hline
B = Most probable expected disease(s) \\
1115 ['Ventral hernia : bulging of the abdominal wall'] \\
825 ['Vocal cord paralysis : voice box weakness'] \\
771 ['Torticollis, spasmodic : neck muscle spasm'] \\
1503 ['Scar'] \\
\hline
\end{tabular}
\end{center}

**Fig. 4.** Given a set of symptoms(A), the predicted(B) and expected outputs(C) are shown respectively.

5 Conclusion and Future scope

Indeed recommender systems play a major role in everybody’s daily life covering online shopping, movie streaming, etc. With state-of-the-art potential of recommender system, this can be extended to healthcare department to aid in a variety of tasks such as managing resources of a healthcare institution, replacement suggestions for equipment(s), recommending medical advice and suggestions, etc. In this paper, the proposed model recommends remedies for the predicted disease(s) followed from the given symptoms by using the generated dataset\(^3\) consisting of a list of most favourable remedies corresponding to a wide range of disease(s). The future scope for this system includes improving the dataset by categorising the mentioned remedies under labels such as: self-care, medication, surgical procedures, non-surgical procedures, therapies, etc., incorporating more filtering algorithms for better results, creating a web based UI for better interaction with the proposed system.

\(^3\) [https://github.com/sud0lancer/Diagonosis-Precaution-dataset](https://github.com/sud0lancer/Diagonosis-Precaution-dataset)
Acknowledgment

This research is supported by “ASEAN- India Science & Technology Development Fund (AISTDF)”, SERB, Sanction letter no. – IMRC/AISTDF/R&D/P-6/2017. Authors are also thankful to the authorities of “Indian Institute of Information Technology, Allahabad at Prayagraj”, for providing us with the infrastructure and necessary support.

References

1. Jong-Hun Kim, Jung-Hyun Lee, Jee-Song Park, Young-Ho Lee, Kee-Wook Rim, (2009, November), Design of Diet Recommendation System for Healthcare Service Based on User Information, 2009 Fourth International Conference on Computer Sciences and Convergence Information Technology (2009). https://doi.org/10.1109/ICCTIT.2009.293

2. A. S. Hussein, W. M. Omar, Xue Li, Modafar Ati, (2012, December), Efficient Chronic Disease Diagnosis prediction and recommendation system, 2012 IEEE-EMBS Conference on Biomedical Engineering and Sciences (2012). https://doi.org/10.1109/IECBES.2012.6498117

3. Yong-Feng Huang, Peng Liu, Qiao Pan, Jing-Sheng Lin, (2012, December), A doctor recommendation algorithm based on doctor performances and patient preferences, 2012 International Conference on Wavelet Active Media Technology and Information Processing (ICWAMTIP) (2012). https://doi.org/10.1109/ICWAMTIP.2012.6413447

4. E. Sezgin, S. Ozkan, (2013, November), A Systematic Literature Review on Health Recommender Systems, 2013 E-Health and Bioengineering Conference (EHB) (2013). https://doi.org/10.1109/EHB.2013.6707219

5. Jhonny Pincay, Luis Terán, Edy Portmann, (April, 2019), Health Recommender Systems: A State-of-the-art Review, 2019 Sixth International Conference on eDemocracy & eGovernment (ICEDEG) (2019). https://doi.org/10.1109/ICEDEG.2019.8734362

6. Madhusree Kuanr, Puspanjali Mohapatra, Jayshree Piri, (January, 2021), Health Recommender System for Cervical Cancer Prognosis in Women, 2021 6th International Conference on Inventive Computation Technologies (ICICT) (2021). https://doi.org/10.1109/ICICT50816.2021.9358540

7. Agarwal, S., Punn, N.S., Sonibhadra, S.K., Tanveer, M., Nagabhushan, P., Pandian, K., Saxena, P.: Unleashing the power of disruptive and emerging technologies amid covid-19: A detailed review. arXiv preprint arXiv:2005.11507 (2020)

8. for Biotechnology Information, N.C.: National center for biotechnology information. https://ncbi.nlm.nih.gov (2021), online; accessed 4 June 2021

9. Cheung, K.L., Durusu, D., Sui, X., de Vries, H.: How recommender systems could support and enhance computer-tailored digital health programs: a scoping review. Digital health 5, 2055207618824727 (2019)

10. Cleavelandclinic: my.cleavelandclinic.org, https://my.cleavelandclinic.org (2021), online; accessed 1 May 2021

11. DermNet: Dermnet nz. https://dermnetnz.org (2021), online; accessed 15 April 2021
33. N. Mohammadi, M.H.B.: N. mohammadi, m. h. babaei, (2014, february), recommending an appropriate doctor to a patient based on fuzzy logic, int. journal of current life sciences, vol.4, issue 2, pp.403-407. Int. Journal of Current Life Science (2014)
34. Nature: Prognostic markers - latest research and news. [https://www.nature.com/subjects/prognostic-markers](https://www.nature.com/subjects/prognostic-markers) (2021), online; accessed 8 October 2020
35. Peito, J.: Incorporating complex domain knowledge into a recommender system in the healthcare sector. Master’s thesis, NSBE: NOVA – School of Business and Economics (2020)
36. Punn, N.S., Agarwal, S.: Multi-modality encoded fusion with 3d inception u-net and decoder model for brain tumor segmentation. Multimedia Tools and Applications pp. 1–16 (2020)
37. Schäfer, H., Hors-Fraile, S., Karumur, R.P., Calero Valdez, A., Said, A., Toklamaan, H., Ulmer, T., Trattner, C.: Towards health (aware) recommender systems. In: Proceedings of the 2017 international conference on digital health. pp. 157–161 (2017)
38. Science, T.D.: Recommendation systems — models and evaluation. [https://towardsdatascience.com/recommendation-systems-models-and-evaluation-8494-4a84f8b8e](https://towardsdatascience.com/recommendation-systems-models-and-evaluation-8494-4a84f8b8e) (2021), online; accessed 4 June 2021
39. Seitz, R.: Understanding tf-idf and bm25. [https://kmwllc.com](https://kmwllc.com) (2020), online; accessed 28 October 2020
40. Sumanth, U., Punn, N.S., Sonbhadra, S.K., Agarwal, S.: Enhanced behavioral cloning based self-driving car using transfer learning. arXiv preprint arXiv:2007.05740 (2020)
41. Underwood, C.: Use cases of recommendation systems in business – current applications and methods. [https://emerj.com/ai-sector-overviews/use-cases-recommendation-systems/](https://emerj.com/ai-sector-overviews/use-cases-recommendation-systems/) (2020), online; accessed 8 October 2020
42. Webmd: Better information, better health. [https://webmd.com](https://webmd.com) (2021), online; accessed 2 May 2021
43. Wikipedia: Biomarker - wikipedia, the free encyclopedia. [https://en.wikipedia.org/wiki/Biomarker](https://en.wikipedia.org/wiki/Biomarker) (2021), online; accessed 8 October 2020
44. Wikipedia: Collaborative filtering - wikipedia, the free encyclopedia. [https://en.wikipedia.org/wiki/Recommender_system](https://en.wikipedia.org/wiki/Recommender_system) (2021), online; accessed 15 December 2020
45. Wikipedia: Prognosis marker - wikipedia, the free encyclopedia. [https://en.wikipedia.org/wiki/Prognosis_marker](https://en.wikipedia.org/wiki/Prognosis_marker) (2021), online; accessed 8 October 2020
46. Wikipedia: Recommender system - wikipedia, the free encyclopedia. [https://en.wikipedia.org/wiki/Recommender_system](https://en.wikipedia.org/wiki/Recommender_system) (2021), online; accessed 8 October 2020