CHAPTER- 6

Chapter – 6 Recommendation Techniques for Health Information Extraction

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Name of the Sub-Title</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1</td>
<td>Introduction</td>
<td>77</td>
</tr>
<tr>
<td>6.2</td>
<td>Related work</td>
<td>79</td>
</tr>
<tr>
<td>6.3</td>
<td>Proposed work</td>
<td>83</td>
</tr>
<tr>
<td>6.3.1</td>
<td>Algorithm</td>
<td>84</td>
</tr>
<tr>
<td>6.3.2</td>
<td>Experimental Evaluation</td>
<td>85</td>
</tr>
<tr>
<td>6.4</td>
<td>Summary</td>
<td>89</td>
</tr>
</tbody>
</table>
Chapter 6

RECOMMENDATION TECHNIQUE FOR HEALTH INFORMATION EXTRACTION

6.1 Introduction

Presently a day, for any sort of data, individuals rely on upon web. They utilize web indexes like Google to pursuit data over web. The inquiries must be precise that will give the data identified with client's Health Care. Be that as it may, there is tremendous measure of data on the web as it's hard to get the significant data effortlessly. If there should arise an occurrence of inquiries on right prescription, individuals for the most part have their own particular inclinations. Wellbeing experts require the adaptability of unstructured content to express their conclusions and treatment systems. In this manner, the vicinity of unstructured content is inescapable, be that as it may, for information mining, data frameworks are obliged to change over the unstructured content to an organized representation for examination. Additionally the human advancement where individuals live will have sway on the quantity of decisions of prescription. So the proposed strategy will be utilized to prescribe best medication and nourishment to the clients or patients. These proposals are in view of the positioning of things and nourishments. The trial results give better exactness contrasted with existing strategies.

The Internet, which brought the most creative change on data society, has likewise brought numerous noteworthy changes of medicinal services administrations. By means of the Internet, getting to data about human services administrations turned out to be moderately simpler for administration purchasers who need satisfactory therapeutic medications. Additionally, purchasers can correspond with specialists to get medicinal advices or to make arrangement by email or moment detachments, which are more advantageous correspondence channels than by telephone. Due to these advantages, a great deal more social insurance administration suppliers began distributed sites for their administration on the Internet aggressively; as an outcome, customers can acquire wide decision of administrations and better administration quality. On the other hand, there are likewise negative impacts created by exponential
development of the social insurance sites. Due to an excess of data accessible, buyers cannot easily pick legitimate social insurance administration among them. Some of them may not have the capacity to judge what medicinal services administrations are useful in light of the fact that assessing those administrations as a rule obliges restorative aptitude. Besides, there may be over-promoting sites that hotshot misrepresented data about administrations. For this situation, social insurance administrations on the Internet may confound administration customers and make them faultier. To help clients to pick a fitting administration among the accessible administrations on the Internet, numerous expediting sites for human services administrations, for example, social insurance web entrances and web indexes have been created [5][6][7][8]. The clients can utilize the handling sites as beginning stages and find fitting medicinal services administrations utilizing them. This change permits the clients to get to data about the administrations much simpler than some time recently, and the social insurance suppliers to spare more lives. The facilitating sites, then again, demonstrated their constraint that more complex system is needed in the area of human services. The greater part of the clients who does not have any learning about medicinal services or any thought what isn't right with their bodies can't figure out legitimate human services administrations. What they need is not sorted out data about administrations, but rather an expert rule to the most suitable administrations for a particular client.

In this manner, suggestion frameworks for better medicinal services are proposed. Proposal framework for the human services is a site that suggests medicinal services benefits or gives helpful data to the clients considering. Medicinal services Provider Recommendation System [3] is an illustration of very much proposed social insurance suggestion framework. Client can look the medicinal services suppliers utilizing area, suppliers' claim to fame, and notoriety. Then again, what this framework couldn't understand yet is that fledgling client’s patients still will most likely be unable to discover fitting treatment for them when they don't have the foggiest idea about their precise wellbeing status. Since the vast majority of individual’s absence of therapeutic information, the framework may not be viable, in actuality. To prescribe fitting social insurance administrations to fledgling or
nonprofessional clients all the more successfully, proposal framework must be mindful of clients' vital settings, for example, area, as well as clients' wellbeing status.

The data on the web has gotten to be testing issue for clients. The client should be clear about their needs to get right data. The measure of web-base data accessible on the web has turned into the most difficult occupation today's situation. Individuals are keen on significant and intrigued data from the web. There is immense measure of data on the web so it's hard to get important data identified with clients Health Care and for this client needs more canny system (agents) to assemble the valuable data. If there should arise an occurrence of pursuits on right nourishment and activity, individuals for the most part have their own particular inclinations. Likewise individuals are limited to some restorative conditions so a few nourishments and activities are dodged so they are pulled in towards other sustenance and activities. Likewise the development where individuals dwell will have sway on the quantity of decisions and mixed bags of nourishment. So there is requirement for simple to utilize edge work for nourishment and activity proposal. This system is to examine client's inclination and will and will construct a sustained and wellbeing related client's profile and will utilize the profile to sort the related information with the goal that clients can make flavorful nourishment and activity request.

6.2 Related Work

Information retrieval is a system that retrieves information from web resources on behalf of user requests. User type something on search engine and the requested query is resolved according to some relevance and ranking algorithms on server end to reply the requested information of that user. The fact that user has to enter his query explicitly has itself became serious problem with information retrieval systems. This is because today tremendous information is available on web out of which user may be unaware about most of it.

User will request only for that information that he or she had heard before this. Other existing information in which user may find interest never comes to the user screen because IR returns only relevant information. But unfortunately IR can’t understand user’s interests intelligently and hence lots of useful information remains
behind the scene. Another crucial point is that though this accepts IR in its current form; user has to enter exact and specific query that represents his or her interest. Otherwise IR may come with large set of generalized results and expected result may be displayed somewhere at the later pages of results which is more cumbersome. Instead of this what if users do not have to enter any query at all and someone intelligently recommend customized information for user. Well web personalization is the area where all such care is taken about users and their likes-dislikes.

Web Personalization is a technique that customizes a website in accordance of user’s interests and likes-dislikes. Web personalization does not need any explicit data instead it collects web data in web context which can be structural, content or user profile data. Web Personalization guides the users to achieve better web experience. Web experience can be simply browsing of a web site or it can be as important as purchasing some products or downloading some items from that web site. And this experience can be enhanced by rearranging the web site content or structure, highlighting web links, insertion of some runtime links or creation of new windows or pages. Uses of proper web mining techniques that are intelligent and computationally efficient are required. Accurate utilization of such techniques will lead to web personalization in true sense. In addition lots of problems associated with web personalization also need to be addressed simultaneously. They are cold-start problem, scalability, adapting to user context, managing dynamics in user interests, robustness, information security.

Recommender system is a simulation of Web Personalization. Recommender system is a special kind of personalized information retrieval system that retrieves or say recommends products/items in accordance of user’s interests and likes-dislikes but without explicit requests from users. They are also seen. Recommender engine is an important part for any such commercial website today. The implicit data obtained from intelligent mining techniques is the input for recommender engine and the algorithm implemented inside do the rating predictions and ranking of information to be recommended. One way to classify recommender systems as content based type, collaborative type and hybrid type [1][2][3]. Another way is heuristic based type and model based type recommender systems.
Lots of different factors are important in deciding the quality of a recommender system. These factors are seen as more or less important depending on the application where you are going to use the recommender system. As stated earlier recommender system is simulation of web personalization so improvements in problems of web personalization will improve the quality of recommender system also. Some of factors are accuracy, diversity, scalability, reliability, serendipity etc.

Precision is the property of recommender framework that chooses whether created proposals are precisely recognized to client's hobbies and preferences loathes or not. If user is receiving what he or she is expecting then the accuracy of your recommender system is high.

Diversity is the property of recommender system that forces to recommend various different items as possible. Many it so happens that user get bored if received recommendations are all of same type and in this situation if he or she found items that stands out differently among other items he or she may go for that outstanding member. Basic reason behind this is the human mind set which is easily attracted to odd figure in search of something new than regular routine.

As in [25] proposed some impromptu systems to rank things for incorporation in suggestion list. As indicated by creator most extreme comparability in target inquiry and cases to be recovered is the general methodology in numerous areas yet it doesn't work in a few spaces.

As in [26] proposed three new calculations for enhancing singular differing qualities. As indicated by creator differing qualities issue is dependably been restriction for substance based suggestion methods and the proposed calculations have framed a benchmark on this worry. Out of these Bounded Greedy Selection calculation has enormously lessened the recovery cost and created negligible loss of likeness among target inquiry and proposals.

As in [27] proposed theme expansion, another heuristic way to deal with streamline the harmony in the middle of precision and differences in order to keep
exactness in a specific level while expanding assorted qualities, particularly for
suggestion records acquired as an after-effect of something based community oriented
separating calculation. Point expansion looks like to Osmotic Pressure similarity
where particular porousness is the key criteria for streamlining. Scientific
categorizations are made for different areas, masterminded hierarchically. Every item
has a place with one or more scientific categorizations and they likewise have content
depictions identifying with this space scientific categorizations. The creators
additionally propose intra-list closeness, another metric which is appropriate to catch
the differing qualities utilizing proposed calculation. As per creators compelling
utilization of substance portrayals alongside significance weights of items has
successful effect while positioning things and that is the place the proposed strategy
contrasts from other existing ones. Their exploratory results demonstrated that clients
favoured the modified broadened list even some loss of precision happened, than the
exact unaltered rundown.

As in [4] indicated how essential outline decisions the result, and in this
manner supervisors can pick recommender plans that are more predictable with their
business objectives and customers' inclinations. They found that recommenders can
expand deals, and recommenders that rebate prominence properly might build deals
more. As in [28] proposed a methodology that tries to discover most ideal subset of
things to be suggested over every single conceivable subset. Here resultant rundown's
likeness to target inquiry and differing qualities inside of rundown these two are taken
as a parallel improvement issue. Another assessment measurement, thing oddity, is
proposed. Thing curiosity implies how much a thing is not quite the same as existing
things list. Thing oddity relies on other existing things in client profile. Thing
curiosity brings certain level of trouble for suggestions and thus can be utilized to
produce helpful experiments. By altering the oddity esteem the resistance in precision
misfortune is adjusted. Creator calls attention to that likelihood of prescribing novel
things is low at whatever point similitude is the essential choice standard.

As in [29] suggested that however different oddity and assorted qualities
measurements are prominent in writing of recommender frameworks, they don't
address two critical attributes thing positioning and pertinence. Creator focused on two ground ideas, specifically thing similitude and client thing communication. Client thing collaboration is demonstrated in light of three conditions decision, revelation and importance. They attempted to cover and sum up the old measurements and put in better configuration.

As in [30] suggested that aim situated data recovery differing qualities can be connected enhancing Recommendation assorted qualities. They formalized the assorted qualities and oddity measurements and their outcomes demonstrated that subsequent broadening systems can give best results. Moreover these proposed measurements are very much aware of positioning and pertinence issues brought up in [6].

As in [31] proposed a methodology that is totally orthogonal to the standard calculations followed in conventional recommender frameworks. The variation they proposed is k furthest neighbours (kFN) calculation that uses an enlarged reversed closeness measure. The minimum preferred things of neighbourhoods are suggested. Their outcomes on standard datasets have demonstrated that differing qualities is enhanced with inconsequential misfortune in proposal exactness.

6.3 Proposed Work

![Health care Recommendation framework](image)

**Figure.6.1 Health care Recommendation framework**

An utilization of clients profile which contain the HealthCare inclinations and the preferences and aversions of the sustenance as it is required for customizing proposals, for example, nourishment, works out, wellbeing condition as shown in
fugure1. From clients profile the application will keep up the historical backdrop of the clients and keeps the record of the data into the databases. At whatever point client will look for pertinent information identified with wellbeing the favoured sustenance and activity would be prescribed to the client via seeking the wellbeing condition from the database that the client would prescribe to the client via seeking the wellbeing form from the record that the client has past any wellbeing related issues. Database is expected to keep up the record of the clients profile in light of Health Care data. This is helpful when the client scans for significant eating regimen and activity this is anything but difficult to perceive from the database that what regards suggest the client prerequisite via looking the wellbeing form. As per the sustenance class’s client can choose the favoured nourishment classification which the client need; additionally the application would prescribe the other related sustenance with having nourishment worth to the clients. Individuals hunt down nourishment and activity for wellbeing condition. The application prescribes sustenance and an activity to the client’s wellbeing condition. Likewise we will prescribe the activity through the client’s classifications. In the proposal Calculation all the data of the patient will be keep up from the data client profile that client has entered. Channels recognizable proof is utilized for what kind of information or inquiry the clients have entered; by condition the execution will be done then it will produce the reports that will prescribe the eating routine, nourishment and activity to the client.

6.3.1 Algorithm

Here, we propose a simple algorithm steps for Healthcare recommendation

**Algorithm:** Algorithm for data set we use WebMD

**Inputs:** Dataset D,

**Outputs:** Recommender items with user feedback

1. Open the website
2. Patient registers his name and all necessary details
3. Create an unique id say Ui
4. Search the patient details in the database db
5. For i = Ui ,search (ui,Db,)
6. If $U_i = db_i$
7. Getdetails($U_i$)
8. **Else** create new id $U_{new}$
9. Identify the patient health conditions say $H_c$
10. Like his checkups and past records Past history (PH).
11. Suggest the necessary medicine based his disease
   a. Daily prerequisite of the nourishment based on the period of the patient.
   b. Food predilections of the client, likes and dislikes
   c. His every day nutritional needs based on age and weight
   d. His reasonable status
   e. Foodstuff allergies
12. Take the feedback $fb$ from the customer along with ratings $fr$
13. For $i=1 \ldots 5$
14. $Fr=\{1,2,3,4,5\}$
15. Update( Item ratings $Ir=fr$ )
16. End procedure.

**6.3.2 Experimental Evaluation**

After For the experiments we have taken the dataset form WebMD [9] website.
This is an online health checker website.

Precision is defined as the division of retrieved documents to relevant documents.

$$\text{precision} = \frac{\text{relevant}(\text{documents}) \cap \text{retrieved}(\text{documents})}{\text{retrieved}(\text{documents})} \ldots \text{eqn}(1)$$

Recall is defined as the fraction of the documents that are successfully retrieved.

$$\text{recall} = \frac{\text{relevant}(\text{documents}) \cap \text{retrieved}(\text{documents})}{\text{relevant}(\text{documents})} \ldots \text{eqn}(2)$$
After For the experiments we have taken the dataset form WebMD [9] website. This is an online health checker website.

Table. 6.1 User ratings

<table>
<thead>
<tr>
<th></th>
<th>USER1</th>
<th>USER2</th>
<th>USER3</th>
<th>USER4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetaminophen</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Aspirin</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Fenoprofen</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Flurbiprofen</td>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diclofenac</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The proposed system uses traditional cosine similarity equation as in eqn(3). Now calculating the similarity distance between User 1 and User2

\[
\text{similarity} = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2} \sqrt{\sum_{i=1}^{n} b_i^2}} \quad \text{...eqn (3)}
\]

\[
\frac{3 \times 5}{\sqrt{3^2 + 2^2 + 5^2} \sqrt{5^2 + 3^2}} = 0.419 ..
\]

Now calculating the distance between User1 and User3

\[
\frac{2 \times 4 + 5 \times 1}{\sqrt{3^2 + 2^2 + 5^2} \sqrt{4^2 + 2^2 + 1}} = 0.4617
\]
Now taking the similarity values from user 1 and user 2, user 1 and user 3 the similarity values are higher to user 1 and user 3 than user 1 and user 2. Hence the user 1 is closer to user 3 than user 2.

### Table 6.2 Example Data Set

<table>
<thead>
<tr>
<th>Name</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetaminophen</td>
<td>Pain relief</td>
</tr>
<tr>
<td>Aspirin</td>
<td>Pain relief</td>
</tr>
<tr>
<td>Fenoprofen</td>
<td>Prevention of tension headaches; migraines; hormone headaches</td>
</tr>
<tr>
<td>Flurbiprofen</td>
<td>Prevention of tension headaches;</td>
</tr>
</tbody>
</table>

Metrics that are used to evaluate performance of proposed system are 1. Precision 2. Recall.

![Fig.6.2 Recall improvement](image)
As shown above figure: 6.2. We tested proposed work under the netbeans environment. We classify the results with weka software. We used core i3 processor and gigabyte mother board and 4 gb ram along with 1tb hard disk. We have taken medicines like Acetaminophen, Aspirin, Fenoprofen, Flurbiprofen, Diclofenac. We have taken them in x axis and putting recall at y axis. Result shows the recall rate of proposed and existing system. It is clearly shown that proposed system works better than existing system. At stage of aspirin existing system recall rate is 0.5 and proposed system recall rate is 0.9. We have another metric as precision.

![Fig.6.3 Precision improvement](image)

As shown above Fig:6.3. We have another metric as precision. We have taken medicines in x axis and putting precision at y axis. Result shows the precision rate of proposed and existing system. It is clearly shown that proposed system works better than existing system. At aspirin existing system recall rate is 0.75 and proposed system recall rate is 0.91.
6.4 Summary

In this thesis, the technique utilized a structure for sustenance and activity proposal. Structure Health Care suggestion framework which will give exact data taking into account clients prerequisite and requirements. This system will utilize semantic web innovation to dissect client's inclinations and will assemble a fed and wellbeing related client's profile and will utilize the profile to order the related learning with the goal that clients can make scrumptious nourishment and activity request. As the structure is completely mechanized the application and methods will be productive, practical and easy to understand that would be less drawn out. In future work, we will on more improvements and accumulate more criticism from clients furthermore we can interface with the human services focuses, which will permit medicinal services suppliers that will help the clients who have restorative issue