Recommendation system based on Collaborative Filtering for resources and educational materials on the Web

Santiago Zapata, Fernanda Lemunguir S.

Abstract—In this work, a study is made to the different types of recommendation systems, the most used algorithms and metrics are described, as well as the problems that arise at the time of their design. The developed system is a Web site, called LibreriaSR, where a recommendation system based on collaborative filtering is implemented, in which the calculation of similarity between users is done using the Pearson correlation coefficient and using the similarity metric, users are obtained more related to the active user in order to calculate the predictions of the content and make recommendations of books that are likely to interest, guided by the idea that if there were users with similar tastes in the past, it is very likely that they have similar tastes in the future.

Keywords—collaborative filtering, recommender system, recommender engine

I. Introduction

At present the great information that can be found stored in the resources and educational materials on the web is very diverse, so it implies that the selection of the appropriate material and identification of the most useful, is restricted to the time of search user. In many cases they can not explore such an amount of alternatives and are forced to choose a material, among those that they have been able to explore, that satisfy their needs in a moderate way or give up their search. Recommendation systems (SR) are part of the solution to the problem of the amount of information that users face on the web. Thus, a system that knows users, that acts as a friend who has read a book or has heard about it and immediately knows that he will like it. Although in social networks or online stores, there are recommendation systems based on explicit feedback, in the case of electronic books there is no system that provides optimal recommendations, based on the analysis of user behavior. In order to adapt the information to each user and make appropriate recommendations, SR will receive information from the client about topics, products or services in which he is interested and the system provides a series of personalized suggestions on a specific topic.

The authors are with the Department of Computer Science, Faculty of Engineering, Technological Metropolitan University, Santiago, CHILE

II. General Background

When making a web application where the recommendation system is in charge of resources and educational materials, it will be possible for the user who is looking for a particular material or book, to have the possibility of broadening his search, since similar material will be suggested and that is within the parameters of the topic that the user is looking for. For example, if you previously described the book "Applied Mathematics to Administration and Economics" by J. Ayra, in a future search you will be recommended a book with similar content that may be "Calculation: Applied to Administration, Economics, Accounting and Social Sciences " by LD Hoffmann, which allows the search for new material, broadening the knowledge on the subject sought and something that is really valued in the present, speeding up the time of exploration on the web on a material and finding everything that is needed in one place. As the user qualifies or searches for material frequently, the application can make more personalized recommendations for each member of the system. The objective will be to recommend to a user those materials that are closest to their preferences obtained explicitly

A. Scope

With the execution benefits users in obtaining materials within a web page, which by means of recommendations will achieve the search for material that is within their interests and frequent searches, which will minimize the search time . This will lead to the development and implementation of a digital content recommendation platform that helps users discover content of interest, with minimal effort and in an efficient manner.

B. Limitations

Disposition by users, in providing information or time, to qualify the topics seen, lack of originality when recommending content affects the user's satisfaction with the recommendation system, when the Recommender System does not have enough information about a user or content, it is very difficult to make recommendations, and above all, valid or accurate recommendations for the user in question. In general, if the contents are not valued, it is not possible to make inference about the interest or taste of the users. In this sense, four aspects are presented in which a recommendation system should help the user: to decide if an object will be of interest, to

compare between several items, to know which will be more interesting for the user and to discover interesting items between a large volume of data and help to explore and find related objects given a content of their interest.

III. THEORETICAL FRAMEWORK

Recommendation Systems (SR) are tools developed to help users find content that may be of interest in large repositories of data, by suggesting items for the user to consult. SR collects user information through a feedback and continuous learning process. It is of interest, to develop a system that helps the user to select documents automatically, avoiding searching the entire volume of books that can be accessed, or at least automatically discriminating items that are not going to interest him. SR is to make the system get to know the user without having to deliver explicit information about tastes and preferences. The SRs make use of an act that exists since the human being has consciousness and intelligence: ask advice or recommendation to experts in the field, follow those individuals who have similar tastes to the user (Collaborative Recommendation Systems) or select objects with characteristics similar to objects that I used before or that are similar to the one I was initially looking for (Recommendation Systems based on Content).

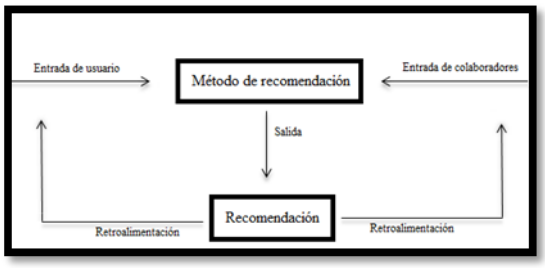

Figure 1: Scheme of the process of generating a recommendation

A. Information Feeding Technique

To achieve the objective of the SR, the tastes and preferences of the users are obtained, the SR provides a mechanism to obtain information on the user's profile, to identify tastes and interest in a specific item, generating personalized recommendations for each member of the community.

1) Explicit Feedback

In this process, users express voluntarily and directly what content they like and seem interesting, through ratings or ratings. The binary scores refer to the explicit interest in an item by the user, for example, I like or do not like it and finally, the scalar scores, which reflect the degree of user compliance with content, for example, Rate from 1 to 5 the material you have seen. Social networks such as Facebook and YouTube use the "like" or "do not like" rating system to rate content.

2) Implicit Feedback

This process consists in evaluating the objects, without the direct intervention of the users, that is, the evaluation is done without the user perceiving it, through the actions that the user performs during the interaction with the system. For example, when the user accesses a material or reads an article online, according to the time it takes to read, the system deduces whether the content is of interest.

B. Classifications of Recomennendation Systems

1) Recommendation Systems based on Collaborative Filtering

It consists of recommending elements that have liked users with similar preferences, based solely on the score they assign to the items of the system (figure 2). Some approaches are based on calculating the similarity between users, and based on their preferences, the elements to be recommended are obtained. The results for the elements not qualified by him, are predicted based on the combination of known scores of close neighbors. Therefore, the recommendation system calculates the similarity between users and creates what they call "close neighbors".

Some advantages are: you do not need a detailed model of preferences, you only need an object valuation vector, you can recommend contents that are difficult to analyze and recommend items based on user preferences. The disadvantages and challenges: Cold Start, Gray Sheep and Sinonimia.

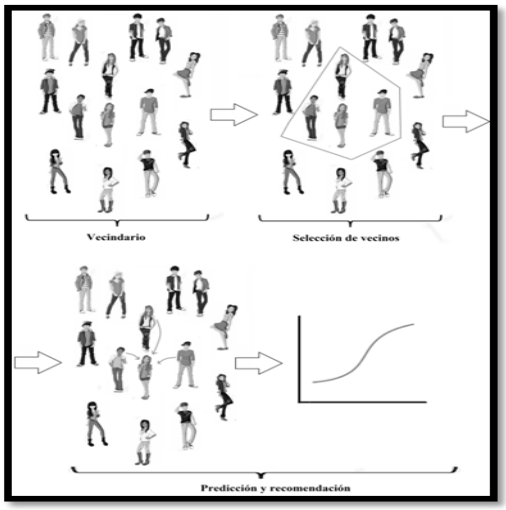

Figure 2: Operation of a Collaborative Filtering Recommendation System
..

2) Recommendation Systems based on Contents

The system allows users to suggest products based on their similarity of content, based on previously liked scores. A preference represents a relationship between a particular user and the data, and more specifically between a user and an attribute associated with the data. In order to know that one item is similar to another, "key words" are searched for the item that the user qualified.

The preferences associated with users can be represented in two ways:

- User profile: It is constructed by selecting the user's preferences, expressed by the user's previous activity or the opinion of the user, collected directly through a questionnaire or through the scoring of an item.

- User model: models the information stored in their profile, to extract some kind of knowledge of this information that defines the user's preferences.

The advantages are: recommendation by content and not by the subjective opinions of other users. The system can generate explanations about the recommendation made based on the user's history.

3) Hybrid Recommendation Systems

These systems use collaborative information and content-based information to generate the recommendations. The classification and development of these systems is complex, due to the wide variety of existing approaches to combine collaborative and content-based information. To create a collaborative content-based hybrid system, user profiles are maintained according to the analysis of the contents of the items, and those profiles are compared directly to determine the similarity among users for a collaborative recommendation.

C. Advantages Recomennendation Systems

The implementation of SR has advantages for users as well as for organizations that use the Web to offer services (music, books, movies, among others). Benefits that help to optimize the search on the Web:

- Search time optimization: search time reduction is generated, increasing satisfaction and increasing interest in discovering new content in users.
- Feedback: A view of the client can be obtained through the feedback mechanism. Helping to know what products are being sold and who is buying them.

D. Problems in the Recomennendation Systems

The implementation of an SR provides benefits, but in its development and design there are problems and challenges.

- Cold start in the SR: when a new user or a new product appears, due to the lack of evaluation of the contents, explicitly or implicitly, the system can not generate appropriate recommendations, since it does not have enough information, which It makes it impossible to carry out the recommendations process. So if you do not value the content, you can not make an inference about the taste or interests of the user.
- Participation of the users: for the system to comply with the objective, the user's participation is needed and it values the contents seen and searched. But users are not always willing to provide personal information about their preferences or spend time assessing or commenting whether or not the content sought is desired.
- Capacity and consumption of resources: the consumption of memory and CPU of any SR is high when dealing with a lot of data. The optimization of algorithms to improve their performance is one of the main fields of research. A constant characteristic of these systems is the processing of constantly modified data, which requires efficient algorithms with a low cost of execution.

A recommendation system requires continuous learning of user profiles and a constant updating of system information.

IV. Proposed Architecture

A. Description of the Problem

With the passage of time, the resources found on the internet grow exponentially, which makes it difficult for users to search for content. In educational resources, such as books, guides, among others, it is a great effort to locate similar material. Searches are limited by the time the user gives to the scan. In this work, there are problems related to the overload of data that are in the Web, Cold Start, costs and mainly to implement efficient information feedback mechanisms, in most cases, the feedback mechanisms are based in the explicit feedback, and this can cause inconveniences to the users because they typically do not like to value the contents.

B. Proposed architecture

The objective is to achieve explicit feedback in SRs of educational materials and resources. So the architecture must constitute three levels:

- First level: User

  In order to recommend interesting content for users, the platform needs to know the behavior of users, the reading environment and the different actions it performs on the contents. This application will serve as a source of information in an educational, professional or leisure environment.

- Second level: Feedback

  The feedback system allows to automatically obtain all the actions carried out by users through the first level. This system allows you to easily and quickly configure the actions that may be interesting to study the behavior of users, either to recommend content related to the profile of users or in order to perform other types of analysis or studies, such as example, statistical analysis, usability study, accessibility evaluation, etc.

- Third level: Explicit data and configuration files

  The explicit data are the result of the analysis and processing of the valuations and the configuration files specify a series of information that is used by the different systems and processes defined in the architecture. In these are configured a series of information necessary for the proper functioning of the platform, how to store in the feedback process, the level of privacy of users.

With the use of the data generated through the explicitation system, the recommendation engine provides the user with a series of contents related to the contents with which he has previously interacted.

C. Explicit Assessment

In order to know the users' interest in a content, explicit feedback is used, an assessment is made from 1 to 5, users rate the contents according to a scale that defines the degree of interest:

- 1: Insufficient
- 2: Regular
- 3: Enough
- 4: Good
- 5: Excellent

D. Programming Language

It is necessary to choose a programming language that fits the needs and requirements. PHP is selected, which is a multiplatform language, oriented to the development of dynamic web applications with access to information stored in a database and has the ability to connect with most of the database engines that are currently used. its connectivity with MySQL and PostgreSQL, has the ability to expand its potential using modules.

V. System Development

The functional requirements in a software describe characteristics that make it practical and utilitarian and that achieve the user's objectives. In general, the properties that the user expects from an SR are the following:

- Rate those elements (books as it is in this project) that you have visited and want to evaluate.
- Receive recommendations, which include objects or items not yet known by the user.
- Edit the user's profile. This functionality supposes:
  - Be able to modify the scores already made.
  - Consult personal data and user preferences.
  - Modify personal data and user preferences.
- Have clear information about all the elements (items) labeled existing in the system, as well as the characteristics of these.

A. Library SR

Librería SR is a network that allows searching and discovering books on the platform, where access to content is facilitated through the web. With the development of this platform, an improvement is sought in the way of interacting with the digital contents and thus adapting them to the needs of each user.

Through the use of this website users can perform a set of activities that will help you discover digital content of interest, such as searching for books and scoring the contents and thus implicitly recommend other users and thereby obtain books that may be of interest

B. Architectural Pattern of Library SR

An architectural pattern is an application design model that offers solutions to software architecture problems. The implemented web application follows an architectural pattern of client-server type. The client initiates data requests, adopting the active role in the communication, that is, it is the master device, while the server waits for the reception of the request, playing the passive role, that is, slave. The use of this pattern in front of others, presents a series of advantages, among which the centralization of control stands out: the server controls both the accesses and the resources used, as well as the integrity of the data, so that a client can not damage the system.

- Interfaces: They are the ones that show the information to the users. Through these the information related to the contents, users and other information of interest for the users is shown. The interfaces are designed using HTML, CSS, JavaScript and Boostrap technologies.
- Application server: This is where operations and calculations reside after receiving requests from users to send a response. This layer communicates with the presentation layer, to receive requests and present the results, and also with the data layer to request the database manager to recover or store data from it. These drivers are developed using the PHP programming language.
- Data Layers: It is responsible for accessing the data stored in the database because it is where the data reside. In this case, the data layer is formed by only one database manager that performs the storage of the data, receives information retrieval requests, as well as the insertion of new data, modification or elimination of them. This data is stored using MySQL as a database manager.

C. Specification of Requeriments

The specification of requirements includes a description of the functionalities, of the behavior, of the system, the functional requirements define the functionalities of the system, and the non-functional requirements: performance, availability, security, accessibility, usability, stability, portability, cost, operability, interoperability, scalability, concurrency, maintainability and system interface.

Functional requirements, the system meets the following requirements in the operation:

- Register users
- Authenticate
- View and modify data
- Rate books
- Recommend books
- Store data
- Close sessions

D. Actors

An actor is an entity external to the system that requests a function and therefore maintains a relationship with it. An actor can be a human being as well as an electronic device and even another software system.

The actors with whom the system interacts are the following

- Unregistered user: The actor is a user not registered in the system. The reason why a user is not registered is because a user is completely new, he has never registered in the system and therefore, the system does not contain information about him.
- Registered user: The actor is a registered user within the system, who has completed the authentication process and can access all system functions.
- Server: The actor is a server that contains the information of the system database. The architecture followed is client-server, where the client is the system that will request or send information to the server.

E. Use Cases

The use cases describe the actions of the system, which produce a determined result and which is of interest to an actor. A use case represents a specific functionality given by the system, what happens and who intervenes. The cases of use are tasks with meaning, coherent and with a certain independence, that the actors perform on a daily basis when using the system. In one use case, one or more actors can participate.

Use Case 1. Unregistered User

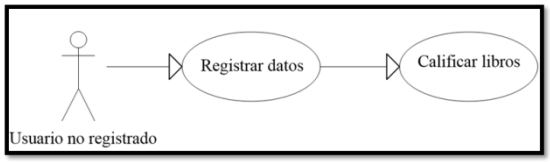

Figure 3: Diagram Case of User use not registered

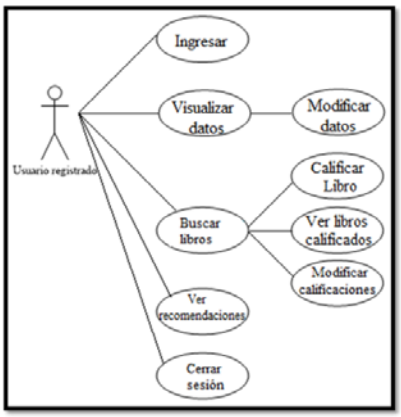

Figure 4: Use Case Diagram for a registered user

F. Database

The database of the web application contains the necessary information so that the application can make use of it and thus provide its service correctly.

Figure 5 is an entity-relationship diagram that contains the structure of the database, determining the identity of the relationship and what type of information is stored within it.

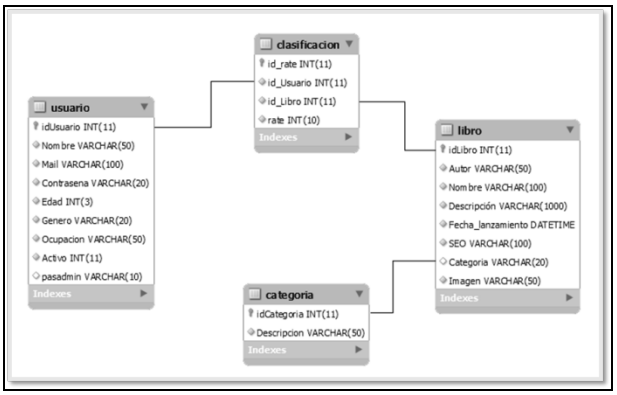

Figure 5: Entity Model - Relationship

VI. Recommendation Engine

A. Model of Mathematical Transformation

To give an assessment that signifies the interest of a user for a book, a set of mathematical operations are defined that with their implementation seek to measure the behavior of the users for each of the actions carried out and with them transform them to a numerical value defined within of a certain range. The established range is defined with the purpose of simulating the explicit valuation of a content, in this system the valuation is from 1 to 5. This means that the lower value would be the lowest rating that the user would give to the content, the higher value would be the maximum value that the user can grant and a value of 0 means that the user has not yet performed an action.

B. Assessment Matrix

The information obtained through the ratings of users to the books available in the system, is represented as a matrix of users and books, where each cell represents the assessment of a user with respect to a specific book.

C. Calculation of Rating for each user

To obtain the Rating of each user that enters the system (active user), all the ratings given to the books that exist in the system are added and then the total is divided by the number of books, that is, an average of Rating for each user. If the sum of all their scores is zero (0), the user has not qualified books and can not obtain a recommendation.

D. Similariti between users

To find k-neighbors, the Pearson correlation (4) is used. The correlation analysis is used to measure the degree of association between two variables dependent on each other. It is measured on a scale that varies between -1 and +1. The value of +1, indicates a perfect and direct correlation, however, the value of -1 indicates that there is an inverse correlation. The correlation value 0, means the absence of correlation between the variables.

The values that are obtained when using the correlation, will indicate the similarity that exists between an active user and a neighbor.

Mathematically, the correlation formula used is:

$$Sim(a,b) = \frac{\sum_{l \in L}\left(r_{a,p} - \overline{r_a}\right)\left(r_{b,p} - \overline{r_b}\right)}{\sqrt{\sum_{l \in L}\left(r_{a,p} - \overline{r_a}\right)^2} \cdot \sqrt{\sum_{l \in L}\left(r_{b,p} - \overline{r_b}\right)^2}}$$

Where:

a,b: Users

l: Book

L: Total books in the system

$r_{a,p}$: User rating "a" to the book "p"

$r_{b,p}$: User rating "b" to the book "p"

$\overline{r_a}$: Average user ratings "a"

$\overline{r_b}$: Average user ratings "b"

Once the similarity values between each neighbor are obtained, you can already choose which of them has greater similarity with the active user.

E. Index of Prediction to books not qualified by a user

To determine the recommendations, a mathematical calculation is made to establish the prediction that a book has towards the user.

This operation uses similarity as a weight and thus weights the scores of the different users in the system.

The formula used for this calculation is

$$predicción(a,l) = \overline{r_a} + \frac{\sum_{b \in N} sim(a,b) \cdot \left(r_{b,p} - \overline{r_b}\right)}{\sum_{b \in N} sim(a,b)}$$

Where:

a: User to recommend.

b: User.

l: Book.

N: Total users in the system.

$r_{b,p}$: Score of user "b" to book "p".

$\overline{r_b}$: Average user ratings "b".

$sim(a,b)$: Similarity between user "a" and user "b".

F. Recommendation Management

According to the number of users of the system, you can configure the recommendation engine and modify the different parameters. This allows the neighbors to increase or decrease, thus modifying the prediction indexes according to the total number of users in the system and with them increasing or decreasing the number of books recommended to the user.

VII. LIBRARY SR

Libreria SR, is a platform that allows access to books with the aim that the user through qualifications, generate recommendations made by means of the similarity that has among the other participants of the system. The purpose of this section is to describe the operations that users can perform on the platform.

A. Main Application Screen

As shown in Figure 6, the user to access the home page, you will see the login to the library in which you can log in, you will also find a link for users who are not registered in the system.

Figure 7: Main page

B. Register as a user

By entering the "Register" link, you access a simple form that allows any visitor to become a user of the platform. In Figure 7, the number of fields needed to encourage the registration of new users has been reduced to the maximum.

Figure 7: Registration

C. User's main page

On the user's homepage, there are two sections, as shown in Figure 8. In the first section are the recommended books and in the second section you can see the latest books added in the library.

Figure 8: User's home page

D. View and rating of a book

As shown in Figure 9, the user can access information in a book, which details the author, description and image of the selected book. On the right side of the screen, you can rate the content seen from 1 to 5.

Figure 9: View and qualification of a book

VIII. CONCLUSIONS

Developing and implementing a Recommendation System for resources and educational materials on the web has been achieved in the proposal, construction and validation of the results of the architecture. The web application contains basic functions to achieve the specifications of the system, these are: registration of users, authentication in the system, rating a book and viewing the recommendations provided by the system. This generates an intuitive interface, the user can visually differentiate the modules arranged on the web. The objective is to know the different algorithms in which the recommendations can be developed, in order to determine which one is appropriate to implement it in the system. The architecture of the SR is in three layers: Interfaces, which are those in which the information is shown to users and they were designed using can be developed, in order to determine which one is appropriate to implement it in the system. The architecture of the SR is in three layers: Interfaces, which are those in which the information is shown to users and they were designed using HTML, CSS, JavaScript and Boostrap. The second layer, Application Server, where the operations and calculations that are executed in the system are nested, these controllers were developed using the PHP programming language, at the end is the data layer, where all the data is stored. system in the database, MySQL was used as a database manager. To implement the recommendation engine, it is based on the collaborative filtering technique

References

[1] Valdez, Edward Rolando Núñez.Sistemas de Recomendación de contenidos para libros inteligentes. Oviedo, ESP., 2012.

[2] McDonal, David, y otros, y otros.PHOAKS: a system for sharing recommendations. USA, ACM1997

[3] Herrera Viedma, Henrique, Porcel, Carlos y Hidalgo, Lorenzo. Internet, Sistemas de recomendaciones: herramientas para el filtrado de información en internet. Hipertext. [En línea] http://www.upf.edu/hipertextnet/numero-2/recomendacion.html.

[4] Pedroza, Henry.Sistema de análisis estadístico con SPSS. Managua, Nicaragua : s.e., 2006.

[5] Verdejo, Vanesssa Gómez. http://www.tsc.uc3m.es/. [En línea] 7 de Noviembre de 2016 http://www.tsc.uc3m.es/~miguel/MLG/adjuntos/VanessaRecommender.pdf.

[6] Francesco Ricci, Lior Rokach, Bracha Shapira. Recommender Systems Handbook. New York. Springer, 2011.

.