International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2016): 79.57 | Impact Factor (2017): 7.296

Hybrid Recommendation System based on Preferences and Consumer Location for the Restaurant Sector of Tlaxcala

Mary Cuecuecha[1], Saúl Pérez[2], José Hernández[3], Federico Ramírez[4]

[1,2]Master in Computer Systems, Apizaco Institute Technological, Apizaco, Tlaxcala, Mexico

[3,4]Departament of Computer and Systems, Apizaco Institute Technological, Apizaco, Tlaxcala, Mexico

Abstract: ***Currently the amount of information that can be accessed through the Internet is colossal, making it difficult to search for products or services that adapt to the requirements (tastes) of each user. For these reasons it has become necessary to build technological tools that provide some kind of reliable suggestions, such as Recommendations Systems, because their main objective is to help users find information about products or services in a better way by filtering all the information available thus achieving a better use of it.In this research work a method of Hybrid Recommendation is presented to create a list of recommended items (restaurants) to users (consumers) of the State of Tlaxcala, merging algorithms: Collaborative Filtering and Content Based. Furthermore, the user experience is improved by applying the user's GPS location as a filter to the recommendations. To measure the performance of the proposed system, we experimented with a set of data extracted from Foursquare and TripAdvisor.***

Keywords: Hybrid Recommendation System, Collaborative Filtering, Content Based Filtering, Natural Language Processing, Bayesian Classification

1. Introduction

The restaurant sector/industry is in a new phase, where it is necessary to use Information and Communications Technology (ICT) that the current globalized world demands to be competitive within to market, representing one of the main sectors in benefit of the economy [1], in addition to promoting tourism in the region, it intends for its consumers to satisfy one of their basic but the most important needs, feeding. In this era of the web there is an overload of information, particularly the search for restaurants that have peculiar characteristics or dishes for a particular consumer is a tedious and often difficult task for this one, because the consumers look for places appropriate to their personal tastes, but due to the large amount of information that can be found on the internet the results obtained are not always the best, which generates a partial satisfaction, being one of the main reasons why recommendation systems play an important role in our daily life [2], currently the problem is being solved by the search engines, but they do not provide the personalization of the data.

Personalization has been recognized as a critical factor for successful restaurant industries and the use of the recommendation systems is the best approach to dealing with the problem of personalization [3].

According to Ricci [4], he says that a Recommendation System (RS) can provide valuable information to assist in the consumer decision-making process with the objective of providing restaurant recommendations with accuracy [3], since the recommendations that are produced in a recommendation system, are reduced to help humans to satisfy their personal tastes and discover new elements, whit less effort, than if they performed the activity manually. In its simplest form, the recommendations are offered as an ordered list of items (for example songs, news or books, in this study are restaurants). When executing this ordering, the Recommendation Systems try to predict which are the most convenient products or services for the user. To achieve this, computationally the Recommendation Systems collect the preferences of the users, either explicitly, for example: asking the users the valuations on the elements, or deducting them from the actions of the users. Currently, Recommendation Systems have proven to be a valuable means of dealing with the problem of information overload [3].

This work presents the bases for the development of a Hybrid Recommendation System according to the classification proposed in the article [5]. Due to use of Artificial Intelligence (AI) techniques, it allows suggesting effective restaurant recommendations, aimed at improving decision making in a simple way and in a considerable time, according to the tastes of each consumer. The motivation of this work is focused on establishing research and development of a recommendation system for the restaurant industry in the region of the State of Tlaxcala. The structure of the work consists of the following form: section 2 related works, section 3 proposed method, section 4 experimental evaluation and section 5 conclusions and future work.

2. Related Works

Initially, recommendation systems emerged as an area of individual research when some researchers began to work on different recommendation problems [6]. The recommendation systems (RSs) have a relation with the systems of search or retrieval of information, since both are designed so that from a set of data relevant information for the user is obtained [7]. But between these systems there are

Licensed Under Creative Commons Attribution CC BY

some specific differences, for example, the user who occupies the search systems must be explicit in the filtering criteria while using them, however in the RSs implicitly obtain the user profile as a result of repeated uses over time. Therefore, in the use of SRs it is important to store certain information that allows generating the user's profile and instead with the search systems its use can be anonymous [8].

Currently the area of the RSs has grown due to the constant evolution of digital platforms that require intelligent technologies to generate knowledge from the large amounts of data that are generated.

In the article [9] it exemplifies Youtube.com, LinkedIn.com and Amazon.com as the main internet sites that operate with a recommendation system. In addition, article [10] proposes a recommendation system for cloud services so that the user can hire the service that meets their own requirements. But the application of the recommendation systems is not only based on the services that exist on the Internet but its popularity of this kind of systems has generated its implementation in different areas of real life.Alvarado, J. Roberto presents in his work [2] a recommendation system for music that applies semi-supervised learning techniques and Guzmán, J. proposes in his work [11] a mobile recommendation system that is based on Intelligence techniques Artificial, allowing a user to suggest an exercise routine.In another case [3] a collaborative recommendation system is outlined using a GIS (Geographic Information System) to suggest tourist products in Cuba.

The RSs have a wide relation with the data mining because the problem of them can be defined as the estimation of lost data. This class of systems includes two fundamental components, the user and the item to be recommended (restaurants, films, videos, books, music, etc.). So the RSsaims to estimate the interest that the user may have aboutthe items to generate the pertinent recommendations from observations and extra information that is obtained from the RSs.

Recommendation Systems are classified into different categories according to the type of information used to recommend products and / or services to users [12].

In the literature you can find a variety of techniques to implement the RSs for example in Figure (1) the most used techniques are shown [8].

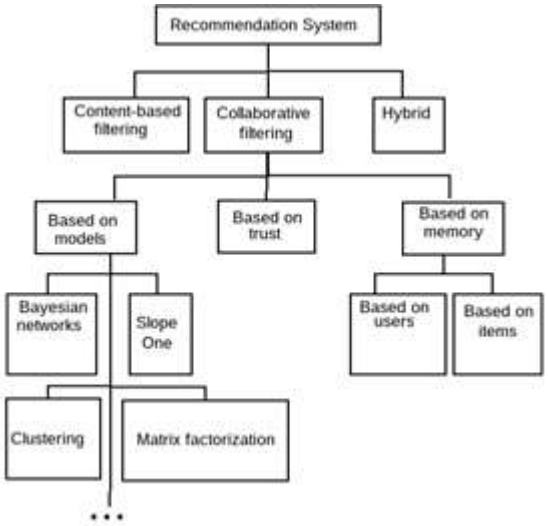

Figure 1: Classification of Recommendation Systems

The main approaches used are collaborative filtering and content-based filtering but both have some limitations and problems, in [12] a review of the recommendation systems is shown which describes the main approaches used in the SRs, in addition to their problems and limitations. So in the majority of the applications that occupy an RS a hybrid approach is used. A hybrid recommendation system combines different techniques to improve the operation of the system. The main idea of the Hybrid Recommender System is to unite different algorithms that generate the recommendations with a better accuracy and efficiency in change when a single algorithm is applied it is possible to obtain inaccurate and unreliable results. Table (1) shows the comparison between the main approaches [6].

Table 1: Comparison of the main approaches of RS

RS approaches	*Number of users*	*Advantages*	*Disadvantages*
Collaborative filtering	Based on several users who have similar interests.	Improve the performance of the recommendation	Data dispersion, scalability, cold start
Content-based filtering	Based on a single user.	Independence of the user, transparency.	Limited content analysis
Hybrid filtering	Combination of Content-Based and Collaborative Filtering.	Reduces the cold start problem and data dispersion.	Degree of high complexity, expensive implementation

Due to the advantages obtained with a Hybrid RS the present work shows the use of the collaborative filtering with the element-based approach and the content-based filtering to generate suggestions for the restaurant sector of Tlaxcala. In the literature there are several applications of a Hybrid RS for example in this article [13] a hybrid recommendation system is proposed to create recommendations on articles with two key components: the collaborative filtering that is incrementally updated and the content based filtering that makes use of latent semantic analysis based on relative term frequency algorithms. Another methodology that takes advantage of collaborative filtering and content-based filtering is presented in this paper [14]. Another case is the

Licensed Under Creative Commons Attribution CC BY

analysis presented by Toine B. [15] where he examines how to incorporate the tags and metadata that are generated in social bookmarking websites by applying the technique of collaborative filtering and content-based filtering. This article [16] shows a RS using a hybrid algorithm to generate personalized recommendations to users during their trip, purchases and other routines that they perform. The use of ontology for the user profile is proposed in [17] where various techniques are used to increase the efficiency of the book recommendation system.And finally in this work [18] a Hybrid RS is proposed to generate music suggestions making use of collaborative filtering and case-based reasoning.

3. Proposed Method

The proposed SR uses a hybrid approach that takes as its main components: users who want to obtain suggestions about the restaurants in the State of Tlaxcala and the items that are represented by the restaurants.

In [19] seven techniques are described to implement a Hybrid RS: Weighted, Switching, Mixed, Feature, Feature augmentation, Cascade and Meta level; in this case Switching is used because the system uses some RS technique depending on the current situation that the user has. For example, if the user does not meet the conditions for using the collaborative filtering, then the content-based filtering is applied. The objective of this hybrid is to meet the needs of different users.

In Figure(2) the flow of the Hybrid Recommendation System is observed, the user initiates session, then it is validated if it has registered qualifications to apply the collaborative filtering, otherwise the user's tasting list that registers at the beginning to use the content-based filtering. When having the list of recommendations, one more filter is applied, taking with reference the current position of the user the closest suggestions are generated.

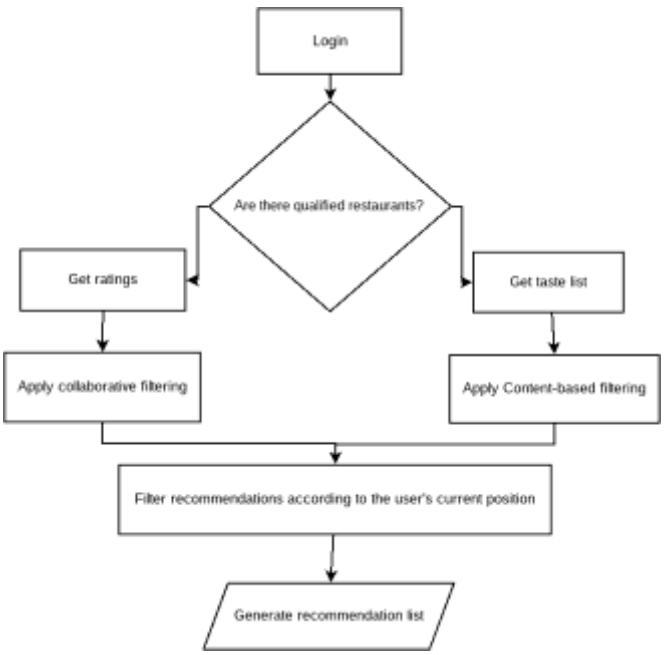

Figure 2: The flow of the Hybrid Recommendation System

3.1 Data Collection

Due to the lack of adequate data for the proposed purposes, the task of constructing a set of data is used to evaluate the persistence of the proposed method. For this, the data is obtained from the Foursquare platform, a technology company that uses smart location to build a meaningful experience for consumer and business solutions [20]. Another platform that deals with is TripAdvisor, a travel site that allows travelers to maximize the potential of each trip by providing content reviews related to travel [21].

The Foursquare site provides an API)[1] that allows you to consume a web service to extract data about places (restaurants, cafeterias, bars, etc.)[2].For this work, the places of the food, drinks and coffee categories [23] found in the State of Tlaxcala were obtained. As a result, 685 places were concentrated with the data described in Table 2, in addition there were 4764 comments related to the places and 65274 registered ratings for 53051 users.

Table 2: Data obtained from the Foursquare API about places

Field	*Description*
id	A unique identifier in string format.
name	Established name.
lat	Latitude value to indicate the location.
lon	Value of the length to indicate the location.
type	Category.

Due to the scarcity of the comments obtained in Foursquare it is decided to use the Web Scraping technique that is used to extract information from web pages in an automated way [24], to implement it the scrapy[3] tool that works with the Python programming language was used. This technique was applied to the site TripAdvisor to get a total of 38107 comments, which were labeled according to the rating given by the user resulting in 14876 comments and with these two classes were created, 7438 are positive and 7438 are negative.

3.2 Description of the Hybrid Recommendation System

(Description HybridRecommendation System)
The main idea of the hybrid recommendation system is to implement the most popular techniques that exist, collaborative and content-based filtering. In addition, the user's location is integrated with the Global Positioning

[1]Application Programming Interface, is a set of subroutines, functions and procedures that offers a certain library to be used by other software as an abstraction layer. [22].

[2]For more information you can visit this website https://developer.foursquare.com

[3]It is an open source tool to extract data from websites quickly, simply and completely. For more information you can visit the site https://scrapy.org

Licensed Under Creative Commons Attribution CC BY

System (GPS) [25] to generate relevant suggestions, because those places that are far from the user are not as important.
In Figure (3) the main components of the hybrid recommendation system are observed.

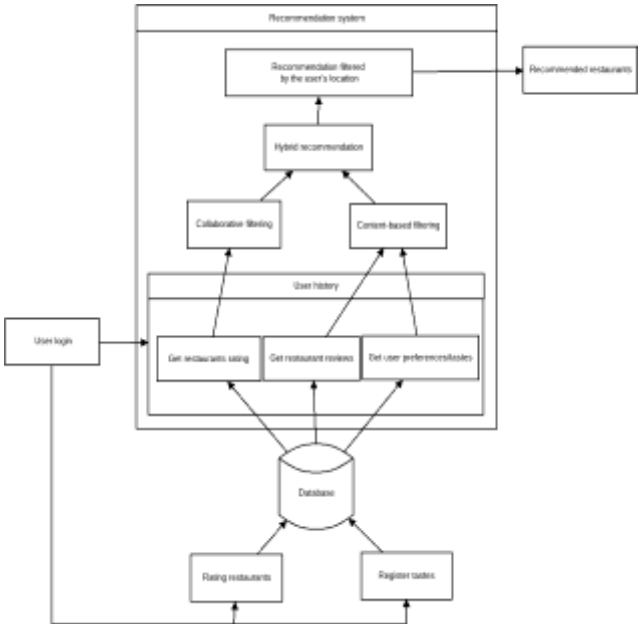

Figure 3: Block diagram of the hybrid recommendation system

The proposal starts when the user registers personal information such as the rating given to each restaurant and the main tastes it has, this is stored in a database for later use.After obtaining the user's history with previously recorded data, of course over time all these data will begin to generate significant recommendations.Then at this point is when the work of the different algorithms that integrate the Hybrid Recommendation System begins.As a first step, the information generated by the useris validated; if the user has registered ratings then all the ratings that were established by him to the restaurants are obtained and used in the collaborative filtering algorithm.

In the article [26] it shows the main collaborative filtering algorithms: user-based and item-based.The proposed recommendation system uses the algorithm based on the item, because the number of users is larger than the number of items and the items do not change frequently.Figure (4) presents the process used by the collaborative filtering algorithm to generate a previous list of recommendations.

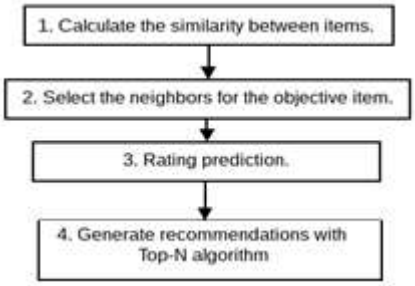

Figure 4: Process of collaborative filtering technique

The first step is to calculate the similarity that exists between the items through the ratings matrix that is created with the information in the database. As an example of the ratings matrix, it is presented in Table (3).

Table 3: Example of a ratings matrix (on a scale of 1-5)

	Item 1	*Item 2*	*Item 3*	*Item 4*
User A	4	?	3	5
User B	?	5	4	?
User C	5	4	2	?
User D	2	4	?	1

For the calculation of similarity, the following measures can be used: pure cosine, adjusted cosine or Pearson correlation coefficient [26]. In this work the Pearson coefficient is applied due to its popularity in the approach based on the item of collaborative filtering. Then the nearest neighbors are chosen for the objective item.With all the neighbors the item qualification is predicted and finally a previous list of recommendations with the Top-N algorithm is generated [27].

In other case when the user does not have ratings restaurants yet, the content-based filtering algorithm is used. Figure(5) shows the algorithm process that generates a previous list of recommendations.

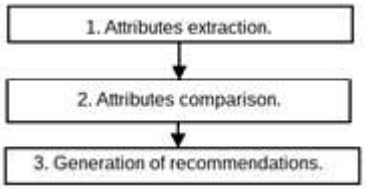

Figure 5: Process of content based filtering technique

The objective of the algorithm is to select and determine the items based on the correlation and relationship between the content of each item (in this work are the restaurants) and the tastes of the users. This algorithm in particular uses text mining techniques: information retrieval, information extraction, categorization and natural language processing [28].To begin the process content-based filtering, the system gets the positive reviews that users have generated for each restaurant and the tastes of the user who wants to get recommendations. The polarity of the comments is measured with a Bayesian classifier [29]. Then with the text mining the attributes of the items are extracted.Later the attributes are compared with the user's tastes and finally the recommendations are generated.
This recommendation system applies one more filter to the recommendations that are generated. With the user's location, the restaurants that are closest to the user are chosen.This completes the flow of the system, generating a list of recommended restaurants for the user.

4. Experimental evaluation

The evaluation of the algorithms integrated in the Hybrid Recommendation System that generates recommendations for users, makes use of a database with data obtained of TripAdvisor and Foursquare.Table (4) describes the data used in the tests.

Licensed Under Creative Commons Attribution CC BY

Table 4: Dataset used in the tests

Dataset	
685	Places like restaurants, bars, cafeterias, among others.
4764	Comments related to places.
65274	Ratings awarded to places.
53051	Users that interact with the places.
14876	Comments used in the Bayesian classifier.

4.1 Accuracy and Recall

The quality of a recommendation system can be evaluated using different types of metrics to evaluate the accuracy or coverage. The type of metric depends on the filtering technique that is used. Accuracy is the fraction of correct recommendations from the total of possible recommendations, while coverage measures the fraction of objects in the search space for which the system can provide recommendations. Metrics to measure the accuracy of a recommendation system are divided into statistics and decision support accuracy [19]. For this study, the decision support precision metrics are used: Precision and Recall, which help users to select very high quality items from the set of available items.

Precision is the fraction of recommended elements that is really relevant for the user [19].

$$Accuracy = \frac{Correctly\ recommended\ items}{Total\ recommended\ items} \quad (1)$$

Recall can be defined as the fraction of relevant elements that are also part of the set of recommended elements [19].

$$Recall = \frac{Correctly\ recommended\ items}{Total\ useful\ recommended\ items} \quad (2)$$

4.2 Evaluation results

After calculating the Precision and Recall values for each user, the **average** of these metrics was obtained for the system using only the collaborative filtering, later the same was applied to the content-based filtering and at the end the calculation was made for the Hybrid Recommendation System. Figures (6, 7) show a comparison of the results for Precision and Recall of the different tests that were performed with the Recommendation System.

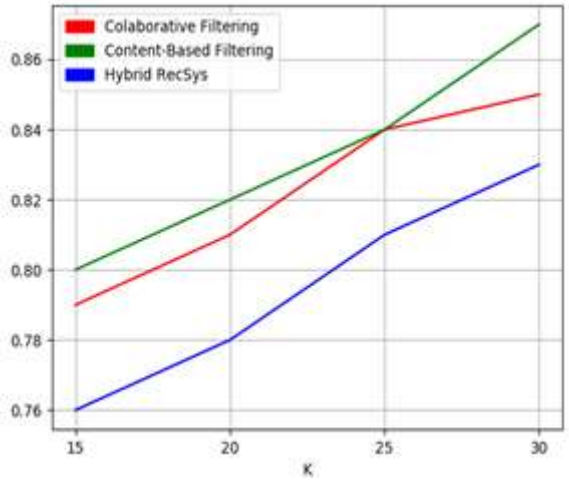

Figure 6: Comparison of Precision between Collaborative Filtering, Content Based Filtering and Hybrid RS

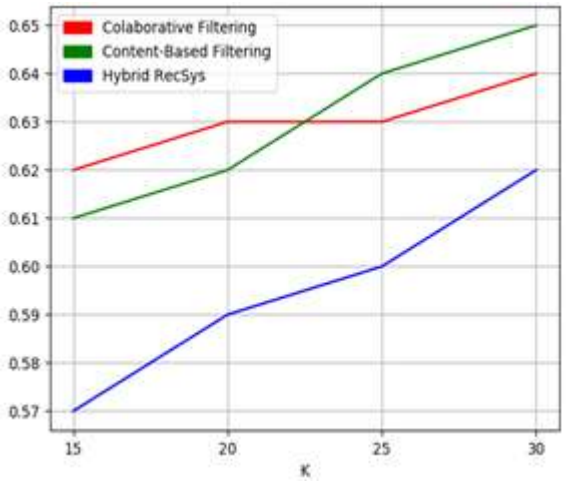

Figure 7: Recall comparison between Collaborative Filtering, Content Based Filtering and Hybrid RS

As can be seen, the results presented in the graph of the increase accuracy increases as the value of k (15, 20, 25, 30)for all cases.The probability that a system-recommended item is relevant to the user exceeds 78% for the 4 variants tested in the case of the Collaborative and Content-Based Filteringand remains above 76% when the Hybrid RS is used.

In the case of Recall, the probability that a relevant item is recommended also increases in direct relation to the value of k, and this probability remains above 60% when evaluating the Collaborative and Content Based Filtering and over 50% for the evaluation of the Hybrid RS.

The decrease in Accuracy and Recall when hybrid SR is used is an acceptable sacrifice considering that, a priori, a significantly higher number of valuation predictions can be calculated in the prediction space, thus achieving a significantly improved coverage higher.

5. Conclusions and Future Work

In this work we have presented a recommendation system that generates a list of restaurants recommended for consumers in the region of the State of Tlaxcala based on the preferences and location of the consumer. The problem was solved by applying techniques of text mining and data mining, using the collaborative and content-based filtering algorithms.

Among the objectives proposed at the beginning of the work were: build a set of data with the theme of restaurants. And on the other hand, evaluate the effectiveness of the hybrid recommendation system with the different cases that may arise when meeting the needs of consumers.

The results obtained showed that the hybrid technique is better than the content-based and collaborative filtering algorithm. Due to the disadvantages that collaborative and content-based filtering has, they are reduced with the hybrid technique. With this we can realize that the combination of algorithms provides a list of recommendations according to

Licensed Under Creative Commons Attribution CC BY

the user.It is important to mention that text mining and data mining are research areas that have a wide application sector. As future work it is proposed to apply a software development technique to implement the Hybrid Recommendation System. For such a situation it is convenient to implement a web service to connect the system with the mobile devices and thus have a further advantage in the system, that the consumer has the recommendations of the restaurants in his mobile device.

References

[1] Instituto Nacional de Estadística y Geografía : Censos económicos 2014. resultados definitivos (Febrero 2018).
[2] García, J.R.A., García, J.V.H., Tello, E.V., de la Rosa, G.R., Sánchez, C.S.: Sistema de recomendación de música basado en aprendizaje semi-supervisado. In: Research in Computing Science.
[3] Otero, J.D.C., Gonzalez, M.J.R., Franco, M.T.M.: Esbozo de una técnica para la recomendación de productos turísticos en cuba usando gis . In: Research in Computing Science.
[4] Stabb, S., Werther, H., Ricci, F., Zipf, A., Gretzel, U., Fesenmaier, D.R., Paris, C., Knoblock, C.: Intelligent systems for tourism. IEEE Intelligent Systems 17(6) (Nov 2002) 53–66.
[5] Burke, R.: The adaptive web. Springer-Verlag, Berlin, Heidelberg (2007) 377–408.
[6] Patel, B., Desai, P., Panchal, U.: Methods of recommender system: A review. In: 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS). (March 2017) 1–4.
[7] Amatriain, X.: Recommender systems (machine learning summer school 2014 @cmu) (Febrero 2018).
[8] González, O.E., Jacques, S.M.: Estado del arte en los sistemas de recomendación. In: Research in Computing Science.
[9] Jain, S., Grover, A., Thakur, P.S., Choudhary, S.K.: Trends, problems and solutions of recommender system. In: International Conference on Computing, Communication Automation. (May 2015) 955–958.
[10] Ashraf, A., Rabbani, I.M., Martinez-Enriquez, A.M., Muhammad, A.: A user-centered approach for cloud service selection and recommendation. In: Research in Computing Science.
[11] Guzmán-Luna, J., Torres, I.D., Vallejo, S.: Un sistema recomendador móvil de rutinas de ejercicio basado en el perfil del usuario. In: Research in Computing Science.
[12] Mansur, F., Patel, V., Patel, M.: A review on recommender systems. In: 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS). (March 2017) 1–6.
[13] Wang, H., Zhang, P., Lu, T., Gu, H., Gu, N.: Hybrid recommendation model based on incremental collaborative filtering and content-based algorithms. In: 2017 IEEE 21st International Conference on Computer Supported Cooperative Work in Design (CSCWD). (April 2017) 337–342.
[14] Pal, A., Parhi, P., Aggarwal, M.: An improved content based collaborative filtering algorithm for movie recommendations. In: 2017 Tenth International Conference on Contemporary Computing (IC3). (Aug 2017) 1–3.
[15] Bogers, T., Van den Bosch, A.: Collaborative and content-based filtering for item recommendation on social bookmarking websites. Submitted to CIKM 9 (2009).
[16] Renjith, S., Anjali, C.: A personalized mobile travel recommender system using hybrid algorithm. In: 2014 First International Conference on Computational Systems and Communications (ICCSC). (Dec 2014) 12–17.
[17] Chandak, M., Girase, S., Mukhopadhyay, D.: Introducing hybrid technique for optimization of book recommender system. Procedia Computer Science 45 (2015) 23 – 31 International Conference on Advanced Computing Technologies and Applications (ICACTA).
[18] Duzen, Z., Aktas, M.S.: An approach to hybrid personalized recommender systems. In: 2016 International Symposium on INnovations in Intelligent SysTems and Applications (INISTA). (Aug 2016) 1–8.
[19] Isinkaye, F., Folajimi, Y., Ojokoh, B.: Recommendation systems: Principles, methods and evaluation. EgyptianInformaticsJournal 16(3) (2015) 261 – 273.
[20] Foursquare, Inc.: Acerca de nosotros (Febrero 2018).
[21] TripAdvisor, Inc.: Acerca de tripadvisor (Febrero 2018).
[22] Fundación Wikimedia, Inc.: Interfaz de programación de aplicaciones (Febrero 2018).
[23] Foursquare, Inc.: Venue categories (Febrero 2018).
[24] Martí, M.: Qué es el web scraping? introducción y herramientas (Febrero 2018).
[25] Fundación Wikimedia, Inc.: Sistema de posicionamiento global (Febrero 2018).
[26] Yang, Z., Wu, B., Zheng, K., Wang, X., Lei, L.: A survey of collaborative filtering-based recommender systems for mobile internet applications. IEEE Access 4 (2016) 3273–3287.
[27] Deshpande, M., Karypis, G.: Item-based top-n recommendation algorithms. ACM Transactions on Information Systems (TOIS) 22(1) (2004) 143–177.
[28] Vijayarani, S., Ilamathi, M.J., Nithya, M.: Preprocessing techniques for text mining-an overview. International Journal of ComputerScience&Communication Networks 5(1) (2015) 7–16.
[29] Sanzón, Y.M., Vilariño, D., Somodevilla, M.J., Zepeda, C., Tovar, M.: Modelos para detectar la polaridad de los mensajes en redes sociales. In: Research in Computing Science

Author Profile

Mary Carmen Cuecuecha Muñoz. She obtained the title of Engineer in Information Technology and Communication in the National Technological Institute of Mexico (TecNM) Apizaco campus, in 2016. Currently is a student of Master in Computer Systems in TecNM campus Apizaco. Her areas of research include: Text Mining, Natural Language Processing.

Saúl Pérez Tirzo. He obtained the title of Engineer in Information Technology and Communication in the National Technological Institute of Mexico (TecNM) Apizaco campus, in 2016. He is a full-stack developer

Licensed Under Creative Commons Attribution CC BY

from 5 years ago. Currently is a student of Master in Computer Systems in TecNM campus Apizaco. His areas of research include: Data Mining, Big Data, Artificial Intelligence.

José Crispín Hernández Hernández. He obtained the title of Lic. In Computing by the Technological Institute of Apizaco (ITA), Apizaco, Tlax., Mexico, in 1993. The degree of M. in C. in Computational Sciences by the ITA, Apizaco, Tlax., Mexico, in 1998. And the degree of Doctorate in Informatique by the Universitéd'Angers, France in 2008. His areas of research include: metaheuristics optimization, bioinformatics, biomedicine, fuzzy logic, computer vision.

José Federico Ramírez Cruz. He is currently a professor at the National Technological Institute of Mexico (TecNM) Apizaco campus. He obtained the degree of Doctor of Science in the area of Computational Sciences in the National Institute of Astrophysics, Optics and Electronics (INAOE) and the Master of Science in Electronics in the same INAOE, and the degree of Industrial Electronics Engineer in the TecNM campus Puebla. His areas of interest are: the applications of Automatic Learning, Evolutionary Computing and Computer Vision.

Licensed Under Creative Commons Attribution CC BY