IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 21, 2019, accepted November 19, 2019, date of publication December 5, 2019,
date of current version December 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2958002

Using Research Literature to Generate Datasets of Implicit Feedback for Recommending Scientific Items

MÁRCIA BARROS[1,2], ANDRÉ MOITINHO[2], AND FRANCISCO M. COUTO[1]
[1]LASIGE, Departamento de Informática, Faculdade de Ciências, Universidade de Lisboa, 1749–016 Lisbon, Portugal
[2]CENTRA, Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, 1749–016 Lisbon, Portugal

Corresponding author: Márcia Barros (mcbarros@fc.ul.pt)

This work was supported by the Fundação para a Ciência e Tecnologia (FCT), under LASIGE Strategic Project - UID/CEC/00408/2019, CENTRA Strategic Project UID/FIS/00099/2019, FCT funded project PTDC/CCI-BIO/28685/2017 and PhD Scholarship SFRH/BD/128840/2017.

ABSTRACT In an age of information overload, we are faced with seemingly endless options from which a small number of choices must be made. For applications such as search engines and online stores, Recommender Systems have long become the key tool for assisting users in their choices. Interestingly, the use of Recommender Systems for recommending scientific items remains a rarity. One difficulty is that the development of such systems depends on the availability of adequate datasets of users' feedback. While there are several datasets available with the ratings of the users for books, music, or films, there is a lack of similar datasets for scientific fields, such as Astronomy and Life and Health Sciences. To address this issue, we propose a methodology that explores scientific literature for generating utility matrices of implicit feedback. The proposed methodology consists in identifying a list of items, finding research articles related to them, extracting the authors from each article, and finally creating a dataset where users are unique authors from the collected articles, and the rating values are the number of articles a unique author wrote about an item. Considering that literature is available for every scientific field, the methodology is in principle applicable to Recommender Systems in any scientific field. The methodology, which we call LIBRETTI (LIterature Based RecommEndaTion of scienTific Items), was assessed in two distinct study cases, Astronomy and Chemistry. Several evaluation metrics for the datasets generated with LIBRETTI were compared to those derived from other available datasets using the same set of recommender algorithms. The results were found to be similar, which provides a solid indication that LIBRETTI is a promising approach for generating datasets of implicit feedback for recommending scientific items.

INDEX TERMS Recommender systems, collaborative filtering, scientific literature, dataset, astronomy, chemical compounds.

I. INTRODUCTION

In the last years, scientific literature has increased in size and complexity [1]. Scientific literature has several applications and purposes, but the main goal is to disseminate the work and the discoveries of researchers. Recommender Systems (RSs) have been a useful help to that end, by improving the discoverability of research articles.

The goal of out article is to provide a methodology for generating datasets of implicit feedback, suitable for evaluating recommender algorithms in scientific areas, by going beyond the recommendation of topics and articles, and support the recommendation of scientific items. For the purposes of this work, we define **scientific item** as an entity belonging to the universe, that may be modeled, characterized by multiple features using a computational representation, and an object of research. Some examples of scientific items are genes, phenotypes, chemical entities, plants, diseases, stars, and groups of stars, such as Open Clusters and Galaxies.

RSs are software tools that provide suggestions for items that are presumably of interest to a particular user [2], which have been used in the recommendation of a wide range of products, for example, movies, books, research articles, or e-commerce [3]–[5]. Some well-known platforms integrating

The associate editor coordinating the review of this manuscript and approving it for publication was Pasquale De Meo.

RSs are GroupLens[1], including MovieLens[2], Amazon[3], Netflix[4], and Google News[5]. Due to the wide applicability of RSs, there has been a progressive interest in the research of new recommendation methods and algorithms. In the beginning the approaches were mostly based in similarity metrics, but now they evolved to machine learning and deep learning techniques [4]–[13].

Recommender algorithms try to predict the interest of the users in each item/product, mostly based on information from their past behaviour. Explicit or implicit feedback from the users may provide this information. Explicit feedback means that the users wittingly indicate if they liked or not some item, for example, by rating an item in a five stars scale. On the contrary, implicit feedback is extracted from the activities of the users, for example, information about what items a user clicked on or purchased. Explicit or implicit information about the preferences of the users is the foundation for RSs, allowing the creation of user/item ratings matrices. Depending on the approach, RSs may be divided into Collaborative-Filtering (CF), when using the similarity between the ratings of the users to provide the recommendations, and Content-Based (CB), when using the similarity between the characteristics of the items, and hybrid, a combination of both CF and CB [14]. CF algorithms may be divided into two methods, memory-based and model-based [10]. Memory-based methods compare users patterns of ratings by calculating the similarity between the rows (users) or the columns (items) of the ratings matrix. Model-based methods use machine learning and data mining to predict the ratings, filling the user/item ratings matrix blank spaces. One of the most used Model-based method is matrix factorization, a method which leverages all row and column correlations in one shot to estimate the entire data matrix [15]. Whereas with Memory-based methods we may explain the recommendations with ''similar users also liked this item'', with Model-based methods it is not always simple to identify the reason why we are recommending an item.

Despite the dissemination of RSs in many fields, for example, movies, music, and e-commerce, they are not being widely used in Science. The main reason is that it is not easy to gather information about the preferences of the users/researchers about an item/topic. Offline evaluation methods [16] for recommender algorithms require a dataset with information about the past interests of the users to compare the ratings that the recommender algorithms predicted with the real ratings. Most of the platforms holding log files about the users have privacy restrictions, keeping these files private and protected.

In Health Sciences there are a few recommender systems that recommended scientific items. Those that exist are mainly focused either on the recommendation of clinical information and research articles to health professionals or on the recommendation of health related content to patients [17], [18]. In addition, drugs, genes, diseases and their relations are also scientific items targeted by recent recommender systems studies [19]–[21]. Other studies focus on the recommendation of plants [22], and nutrition [23]. A common complaint in all studies is the lack of datasets for evaluating recommender systems.

Offline evaluation is suitable for measuring the accuracy of the predicted ratings, for example through Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), and the accuracy of ranked lists of recommended items, for instance, through Precision (PRE), Recall (REC), F-measure (F1) and normalized Discounted Cumulative Gain (nDCG) [24]. MAE measures the difference between the value of the real rating of an item, and the value of the rating predicted by a recommender algorithm, for all n items under analysis. The lower this value, the better the algorithm. For evaluating the predicted rankings the most used metrics are Precision, Recall, and F-measure. The values range between zero and one, and the algorithm is better if it achieves values closest to one. For a given number k of recommended items, Precision is defined as the percentage of recommended items that are relevant for the user. The Recall is the percentage of the total relevant items for a user that has been recommended. For example, if a list of size 10 recommends 5 relevant items for a user whose total number of relevant items on that test set is 5, the Recall will be 100%, because the algorithm is recommending all the possible items the user was interested in. The F-measure is the harmonic mean of Precision and Recall, allowing the global evaluation of the recommender algorithm. The nDCG measure evaluates the quality of the raking. Higher rated items should appear first in the ranking. Offline evaluation requires the division of the dataset into a training set, used for training the system, and test set used for evaluating the system. This information will enable us to compare the rating predicted by the recommender method, with the real rating in the test set.

In most of the scientific and medical fields, evaluation datasets are unavailable, compromising the evaluation and application of RSs. [25] acknowledged the problem above and proposed a solution. They created a dataset (SD4AI) suitable for testing and evaluating RSs for scientific topics by scanning scientific literature for information. This dataset is about the topic of Artificial Intelligence. It consists of 14,143 articles (the articles represent the users in a traditional RS), 18,502 topics related to Artificial Intelligence (which represent the items of a RS) and 1,389,094 ratings. The ratings are the relevance of the topic in the article. This dataset is used to recommend scientific topics and articles using a CF approach.

The goal of our work is to recommend specific items enclosed in the articles. To this end, we develop a methodology, we shall call LIBRETTI - LIterature Based RecommEndaTion of scienTific Items -, based on collecting information from research articles, which are a common artifact in all

[1]http://grouplens.org
[2]http://grouplens.org/datasets/movielens/
[3]http://www.amazon.com
[4]http://www.netflix.com
[5]http://news.google.com

scientific fields. Our approach generates a <user,item,rating> dataset, where authors of research articles represent the users, and the scientific items they wrote about represent the items to recommend. The number of articles an author wrote about an item are the implicit ratings. These ratings represent the strength of the interest of an author for an item. The structure of the dataset is the same as in [25], however, the meaning of user, item and rating is significantly different. [25] recommend topics and articles based on topics and based on articles, whereas with LIBRETTI we are able to recommend scientific items (not topics) to real people (not articles) based on the interests of their peers.

Two interesting fields for testing our approach are Astronomy and Chemistry, because there are well defined lists of scientific items, and it is easy to find research articles related with each item using web services. In the case study of Astronomy, the list of items are open star clusters (in short, Open Clusters or OCs) [26]. The web services used are Simbad[6] [27] and SAO/NASA Astrophysics Data System (ADS)[7] [28]. Simbad is a database of astronomical objects, and ADS is a bibliographic system dedicated to Astronomy. For the case study in Chemistry, the items are Chemical Compounds (Chem) collected from the Chemical Entities of Biological Interest (ChEBI) [29]. This database also includes information about the articles related to each entity, providing the PubMed IDs of these articles. PubMed is a biomedical bibliographic system, and through its web service[8] it is possible to collect the meta-data of each article (e.g.: title, authors, year).

Our methodology is suitable for any scientific field provided there is a list of scientific items and there are research articles related to each item.

The main contributions of our work are:

1) A new methodology (LIBRETTI) to create datasets of implicit feedback through scientific literature, helping researchers to find scientific items of interest. The methodology is designed to be general, in principle applicable to any scientific field;
2) A novel dataset in the field of Astronomy for recommending Open Clusters of stars;
3) A novel dataset in the field of Chemistry for recommending Chemical Compounds.

In this article, we describe the creation of datasets for recommender algorithms using LIBRETTI and present a well-founded study of how such datasets behave with CF algorithms. By applying known and tested recommender algorithms to our datasets, we compare our results with the results obtained for other public datasets: SD4AI and Movielens 100k (ML-100k).

We performed the evaluation of the datasets using the methods implemented in the Collaborative Filtering for Java (CF4J) library [30], which was designed for CF research experiments. Although its main function consists in testing new recommender algorithms, we used the algorithms offered in CF4J to evaluate how they perform with the datasets generated by our work.

The Python implementation of the LIBRETTI methodology for both case studies are available at https://github.com/lasigeBioTM/cARM and at https://github.com/lasigeBioTM/CheRM, as well as the full datasets used in this study.

The rest of this article is organized as follows: Section II describes the work related to recommender systems for research articles, Biomedicine, and Astronomy, showing the lack of RSs for these fields, Section III describes the proposed methodology, Section IV presents the results of this study, Section V provides a discussion of the results, and Section VI draws conclusions and suggests future work.

II. BACKGROUND

RSs have been widely used to recommend items such as movies [31]–[34], music [31], [35], or books [31], [36], i.e., items that in one way or another will benefit the owner of the platform where the RS is implemented. RSs have also been used to recommend scientific articles. [5] surveyed more than 200 articles about RSs for research literature, throughout 16 years. According to the authors, most of RSs for scientific literature were applied to books, education, academic alert services, expert search, venue recommendations, educational events, patents, and even plagiarism detection. This survey concluded that CB had been the most used approach to provide the recommendations in the field of RS for research articles, with most of the RSs using implicit ratings due to the lack of explicit ratings. However, the survey does not present any work whose goal was to recommend scientific items besides documents, neither the use of authors as users of a RS.

In scientific fields, the use of RSs is spreading. Table 1 shows in greater detail important research studies from the biomedical field using RSs. It provides information about the field, what is being considered as users and items, the recommendation approach, if the dataset is considered public and its availability (if it is possible to download and use the dataset). A closer analysis shows us that the interest in RSs have been growing in these fields, CF is the most used approach, and the most tested field is health in general. Only few of the datasets used in the research studies presented in Table 1 are public and available. In Biomedicine, the recommendation of Chemical Compounds does not seem to be a common practice. We have only two examples ([37], [38]). In [37], the authors use CF techniques for recommending Free-Wilson-like fragment to Chemical Compounds. The dataset is not public nor available. [38] aimed at discovering new inorganic compounds from all chemical combinations, using CB methods. The dataset is publically, even though it is not a dataset with <user,item,rating> format.

In another scientific field, Astronomy, there are recent studies with the goal of recommending research articles. For example, ADS implemented on its improved platform a

[6] http://simbad.u-strasbg.fr/simbad/
[7] https://ui.adsabs.harvard.edu/#
[8] https://www.ncbi.nlm.nih.gov/home/develop/api/

TABLE 1. Background studies about the use of recommender systems in bio-medicine, collected from Pubmed.

Year	Article	Field	Users	Items	Approach	Pub/Not pub	Availability
2003	[39]	Genetics	Query of genes	Genes	CF	Public	Not available
2007	[40]	Genetics	Pathway	Genes	CF	Public	Not available
2011	[41]	Health	Clinical hematologists	Leukemia types	CF	Not public	Not available
2011	[42]	Drugs	Chemists	Reagents	CF	Not public	Not available
2013	[43]	Genetics	Bacteria and archaeal type strains	16S rRNA gene sequences	CB	Public	Not available
2014	[44]	Health	Personalized Health Record System users	Personalized health information artifacts	CB	Not public	Not available
2015	[37]	Chemical compounds	Chemical compounds	Free-Wilson-like fragment	CF	Not public	Not available
2015	[45]	Health	Healthcare professionals	Clinical orders (e.g., labs, imaging, medications)	CF	Not public	Not available
2016	[46]	Health	Healthcare professionals	Surveillance levels and scientific literature	CB	Not public	Not available
2016	[47]	Health	Patients	Clinical features	CF	Public	Available
2016	[48]	Genetics	RNA binding proteins	RNA targets	CF	Public	Available
2017	[49]	Health	Cardiac patients	Disease prediction and medical recommendations	Hybrid	Not public	Not available
2017	[50]	Health	Health consumers	Health educational websites from MedlinePlus	CB	Public	Available
2017	[51]	Health	Patients	Therapy	CF	Not public	Not available
2017	[52]	Health	Healthcare professionals	clinical orders (e.g., labs, imaging, medications)	CF	Not public	Not available
2017	[53]	Nutrition	Patients	Meals	CB	Not public	Not available
2018	[20]	Drugs	Drug	Disease	CF	Public	Not available
2018	[38]	Chemical compounds	Inorganic compounds	Chemical relevant compositions	CB	Public	Available
2018	[21]	Drugs	Celllines/patients	Drug responses	CF	Public	Available
2018	[54]	Health	Patients with gestational diabetes mellitus	Blood glucose control	CF	Not public	Not available
2018	[55]	Nutrition	Patients	Nutritional advises	Hybrid	Not public	Not available
2018	[56]	Health	Patients	Treatments	CF	Public	Available
2018	[57]	Health	Patients	Insulin	CB	Public	Not available

service that recommends articles related to the one the user is currently reading, however they do not provide information about the recommender algorithms used [58]. [59] is another example of a system recommending astronomical articles. The author developed a tool that finds similar articles based only on text content from an input article (CB algorithm). The dataset used to develop the tool was collected from ArXiv.[9] The author argues that this tool works robustly, finding relevant articles that are not discovered by other platforms via citations, references or suggestions from ADS. However, they do not provide any quantitative evaluation measure for the system, providing only isolated examples, and without information about the ratings of the users. [60] implemented a different approach, by recommending opinions of other users, instead of an item/object. The authors developed a RS for astronomical observatories, that when a user introduces a query related to an instrument, the system recommends logs written by other researchers, providing positive and negative feedback. Reporting the negative feedback allows that new researches do not make the same mistakes as others. To test the system they used an open source logbook data from the Laser Interferometric Gravitational Observatory (LIGO). The performance of the system was tested using six months of logbooks, by comparing the retrieved logbooks with actual relevant entries, with the system retrieving most of the entries correctly. Despite the promising results, the authors do not present the results using standard metrics, such as Precision and Recall, and neither provide a baseline for comparison, for example, how a random recommender would perform in the test set.

More recently, [25] approached the lack of evaluation datasets for recommender algorithms in Science using a solution based on scientific literature. Their approach consists in extracting the main research topics from a dataset of articles, creating a dataset of <article, topic, cardinality>, where the cardinality is the weight of the topic in the article. This dataset is equivalent to a dataset of <user, item, rating>. The goal is to recommend topics related to the articles, and articles related to each topic. One of the contributions of that work was an evaluation dataset in the field of Artificial Intelligence (SD4AI).

Our proposal goes a step further in the RSs field, mitigating the lack of datasets. Unlike previous works, LIBRETTI recommends not the research articles themselves, but the objects and items mentioned in the articles, such as clusters of stars, Chemical Compounds, diseases. The set of items

[9]https://arxiv.org/

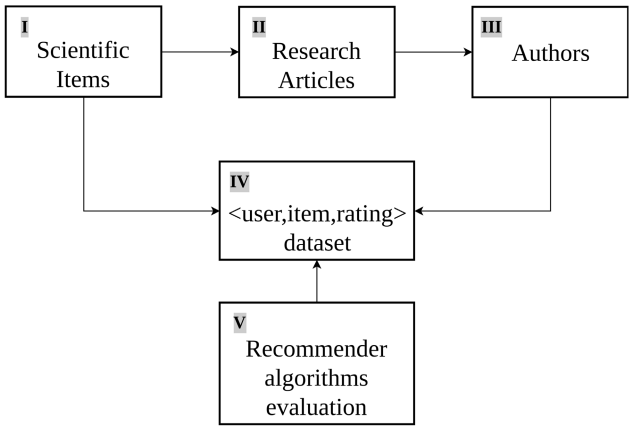


FIGURE 1. General view of the methodology LIBRETTI for creating an evaluation dataset for scientific fields using the scientific literature to extract the implicit ratings.

depends on the scientific field, but as long as they are mentioned or linked to scientific articles our methodology can deal with them. Besides the methodology presented, we also generated datasets for recommending Open Clusters of Stars, and Chemical Compounds.

III. METHODOLOGY

The general view of our methodology, LIBRETTI, for creating datasets for recommending scientific items is represented in Figure 1. The pipeline is as follows:

I. Identification of a list of scientific items;
II. Identification of a corpus of research articles related to each item. This may be achieved by using Named Entity Recognition (NER) to identify the items in the articles, or by using external sources of knowledge, such as Pubmed or ADS, where there is already structured information linking the item to the article;
III. Extraction of the authors from each article;
IV. Generation of the <user,item,rating> dataset. The users are unique authors from the articles, and the rating values are the number of articles a unique author wrote about an item;
V. Evaluation of recommender algorithms using the dataset.

This methodology can be employed in any scientific field with well-defined items, and a corpus where they are mentioned. The next section describes the consolidation of the methodology in Astronomy and Chemistry.

A. STUDY CASES

For testing LIBRETTI we used information from two fields: Astronomy and Chemistry. The consolidation of the methodology for each field is described in the next sections.

1) ASTRONOMY

For the case study using astronomical data, we selected a list of objects from a Catalogue of Open Clusters [26], with 2166 OCs and 13 features. OCs are assortments of stars

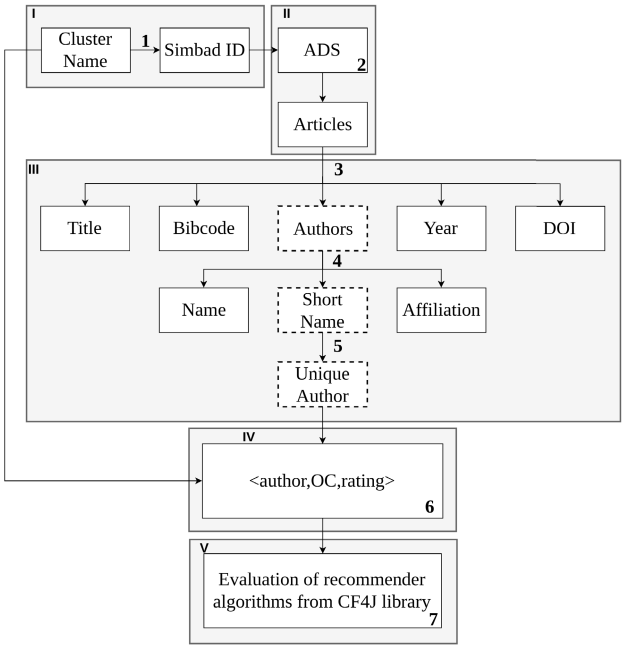


FIGURE 2. Specification of the general methodology described in Figure 1 for a case study in Astronomy, using as scientific items open clusters of stars (OCs).

formed from the same molecular cloud and with approximately the same age. Some attributes of these OCs are the position (galactic latitude and longitude), Diameter, Distance, Age, and Name.

To achieve a <user,item,rating> dataset, where users are authors of scientific research articles and the items are OCs, we followed the steps described bellow (see Figure 2):

1) For each cluster attribute "Name", we searched the unique Simbad ID (unique identifier used by Simbad for each object);
2) Through Simbad ID, using ADS API,[10] we searched all the articles for each cluster, between the years of 1998 and May 2018;
3) For each paper, we extracted the authors, title, year, DOI and bibcode (unique identifier of an article);
4) For each author, we extracted the Name, Short Name, and Affiliation;
5) Next, we identified the unique authors;
6) Finally, we counted how many articles each unique author wrote about each Open Cluster of our list;
7) In this step we used the recommender algorithms provided by the CF4J library with the dataset created in the previous step to access the accuracy of the predicted ratings and accuracy of the given recommendations.

Step 1 required text processing to correct the names of 649 clusters from the catalogue because they were not suitable for searching on Simbad, i.e., searching the clusters by name was not retrieving any Simbad ID. In this regard, it was necessary to identify the non-matching names (usually

[10]https://github.com/adsabs/adsabs-dev-api

due to using an alternative designations) and to correct the spelling. That was done by gathering all names that retrieved null in the first search in step 1, and by finding patterns in the first part of the name. For example, we found 107 names beginning with ASCC + a number. We corrected all these 107 names to [KPR2005] + a number. This happens because the ambiguity of the names: some clusters may have more than one name and not all synonyms are in Simbad. For step 2, before storing the information of the article (authors, title, year, DOI and bibcode), the methodology searches the database for similar bibcodes. If the bibcode already exists, the article is not introduced in the database, storing only the information that this article is also related to the OC under analysis. Step 5 identified the unique authors by finding all authors with the same ShortName and by considering this ShortName as a unique author/user. Step 6 created the <user,item,rating> dataset by counting how many articles a unique author wrote about each OC. The result of this step was a dataset for recommender algorithms for Astronomical OCs (Astronomical Ratings Matrix - ARM).

2) CHEMISTRY

For the case study in Chemistry, the items are Chemical Compounds extracted from ChEBI. Figure 3 shows the steps followed for creating a dataset for recommending Chemical Compounds:

1) From the ChEBI database, we selected all the compounds with 3 stars. For each ChEBI ID, we extracted the PubMed IDs for the articles that are identified in ChEBI as related to that compound;
2) For each PubMed ID, we extracted the information for each article throught PubMed API[11];
3) For each article, we extracted the authors, title, year and DOI;
4) For each author, we extracted the Name. The steps 5, 6, and 7 are the same as in the Astronomical case study, which allows us to create CheRM - ChEBI Ratings Matrix, a dataset for the recommendation of Chemical Compounds.

The correspondence between the general methodology (Figure 1) and its application to the study cases of Astronomy (Figure 2) and Chemistry (Figure 3) is I - 1; II - 2; III - 3, 4, 5; IV - 6; V - 7.

Besides the full ARM dataset and CheRM, we created a subset of ARM and a subset of CheRM by removing all the users with less than 20 rated items (ARM-20 and CheRM-20), to mimic Movielens datasets, where users are only included if they have 20 or more rated items [61]. For these study cases, there was no need to apply NER or any elaborated text-mining techniques since we already have external sources of knowledge with structured information that link the items and the articles. However, in the future, we intend to use these techniques to extract the items and information about them directly from the text of scientific articles.

[11] https://www.ncbi.nlm.nih.gov/home/develop/api/

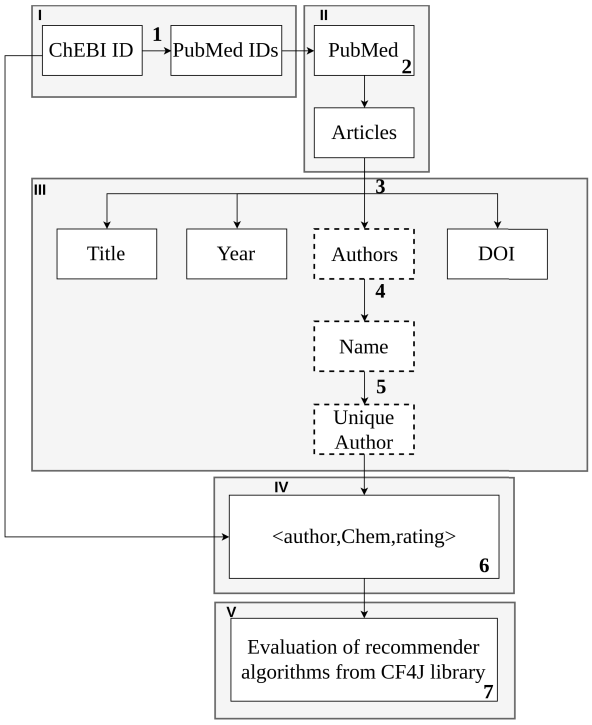


FIGURE 3. Specification of the general methodology described in Figure 1 for a case study in Chemistry, using as scientific items chemical compounds.

B. EVALUATION SETUP

For testing if the ARM and CheRM datasets built with LIBRETTI (see Figure 2 and Figure 3) are suitable for recommending scientific items (Open Clusters and Chemical Compounds, respectively), we followed the setup described below. We applied the same setup to other datasets, namely SD4AI [25] and the dataset from Movielens with 100k ratings (ML-100k) [61]. By following the next steps (Figure 4), this study is entirely replicable.

1) Selection of the evaluation framework. Several libraries exist for evaluating recommender algorithms such as LensKit [62], CF4J [30], and Mahout [63]. In this work we adopt CF4J for the evaluation of our dataset for its simplicity of use and for providing well tested recommender algorithms. CF4J also allows to directly compare our results with the results obtained in [25].
2) Selection of the recommendation methods. CF4J provides a wide range of CF recommender methods, from both memory-based and model-based methods. For this work we selected a k-nearest neighbors algorithm (a memory-based method), with the following similarity metrics: Pearson correlation (COR), Cosine similarity (COS), Proximity-Impact-Popularity (PIP), Jaccard Mean Squared-Difference (JMSD) [64], Jaccard Index (JAC), Mean Squared Differences (MSD). For model-based method, we selected a matrix factorization algorithm, the Probabilistic Matrix Factorization (PMF). With these methods we achieve a wide representation of CF algorithms.

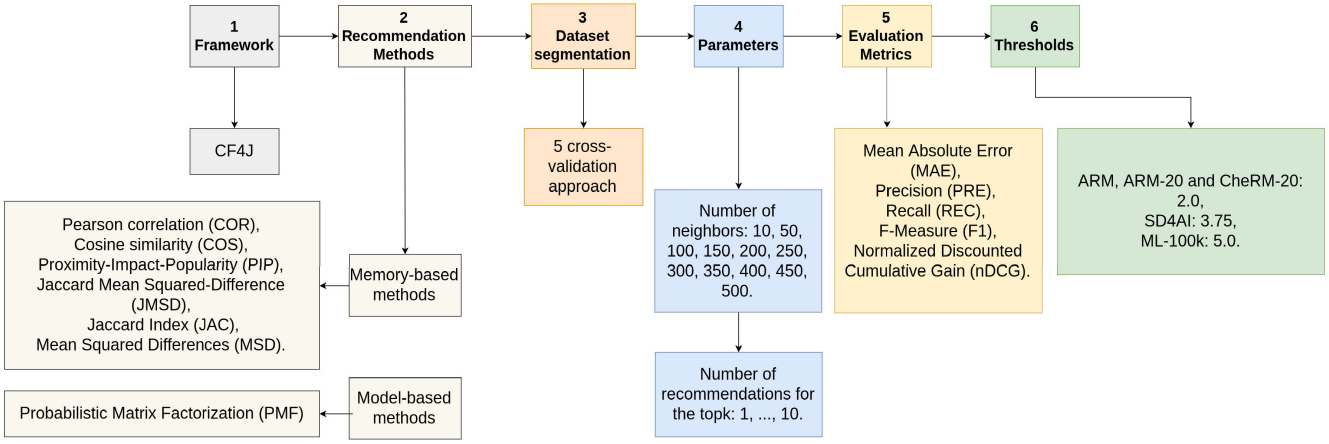


FIGURE 4. Evaluation setup.

TABLE 2. Parameters used in the PMF algorithm for the ML-100k, ARM-20, CheRM-20 and SD4AI datasets.

Dataset	Latent Factor	Iterations
ML-100k	1	50
ARM-20	3	50
CheRM-20	1	150
SD4AI	16	150

3) Segmentation of the dataset for training and testing. In this step we selected a 5 cross-validation approach (20% for the test set and 80% for the training set).
4) Selection of the cross-validation parameters:
   a) Number of neighbors for Memory-based methods: 10, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500;
   b) Number of recommendations for the topk: 1, ..., 10.
   c) For the PMF recommender algorithm the parameters used are described in Table 2. These are the optimal conditions achieved by testing different values.
5) Selection of the evaluation metrics. The algorithms were evaluated for MAE, PRE, REC, F1, and nDCG. With these metrics we evaluate the accuracy of the predicted ratings, the relevance of the recommended items, and the quality of the recommended rankings.
6) Selection of thresholds (minimal rating value for considering a recommended item as relevant for the user, used in the calculation of the Precision, recall and f-measure) for the different datasets being tested. ARM, ARM-20, CheRM-20: threshold 2.0; SD4AI: threshold 3.75; Ml-100k: threshold 5.0.

IV. RESULTS

In this section we describe the results obtained from the application of LIBRETTI to the Astronomy and Chemistry use cases, and the performance of the algorithms in the different datasets.

A. DATASET DESCRIPTION

Following LIBRETTI (Figure 1) applied to the astronomical case study described in Section III-A (Figure 2), we created a database with 2,166 items, 12,378 articles, and 83,208 authors, resulting in 17,006 unique authors, when grouped by equal ShortName. From the 2,166 items, 64 were excluded because no Simbad ID was found. The dataset created from our database has a size of 17,006 rows × 2,102 columns, with 179,269 ratings, which means that our user/item ratings matrix has a level of sparsity of 99.5%. The sparsity level matches the sparsity levels of rating matrices presented by other studies [25], [65], [66]. For the Chemistry case study, we have 22,307 Chemical Compounds (with distinct ChEBI ID), 66,655 articles and 345,494 authors. The final dataset of <Author,Chem,Rating> has 22,299 Chemical Compounds, 193,106 unique authors and 456,681 ratings.

Table 3 shows the dimensions and statistics about the datasets of ARM, ARM-20, CheRM, CheRM-20 and also for SD4AI and ML-100k.

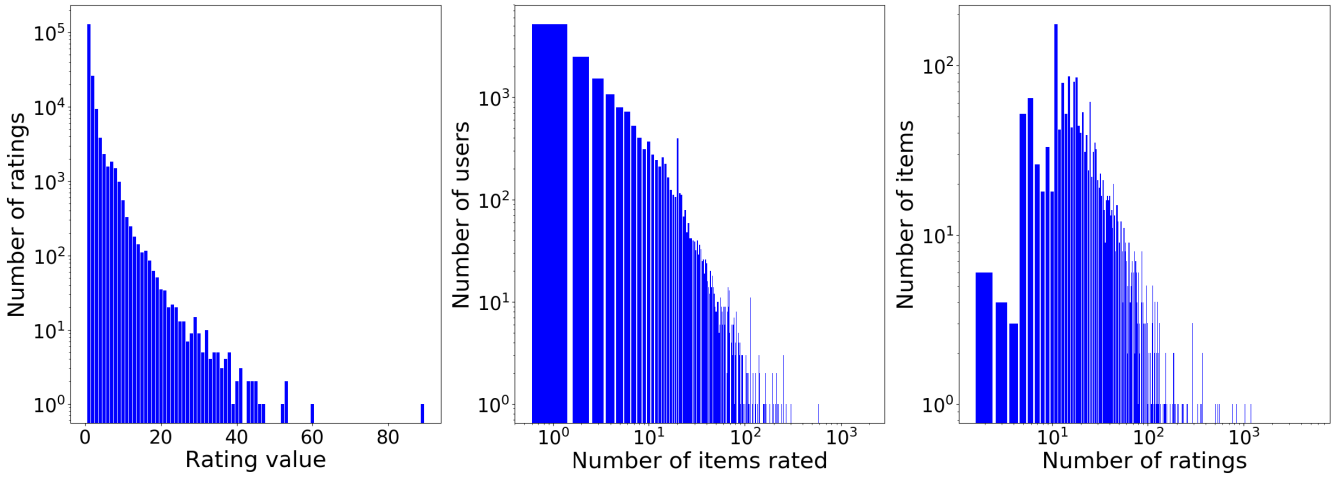
Figure 5 shows the relevant statistical information of ARM (Figure 5a) and CheRM (Figure 5b) datasets. The maximum rating value for ARM is 89 (a single author wrote 89 articles featuring a cluster), corresponding to user 14308 and item "Melotte 22" also known as the Pleiades (simbad ID:675533). For CheRM que maximum rating is 62, corresponding to user 164989 and to the item ChEBI:101096 (ethoxzolamide).

The distribution of the rating values by number of ratings is represented on the left graphics of Figures 5a and 5b. The minimal rating for both datasets is 1, and it corresponds to 72% of the ratings for ARM and 93% for CheRM, meaning that the majority of the authors wrote only one article about the items in study.

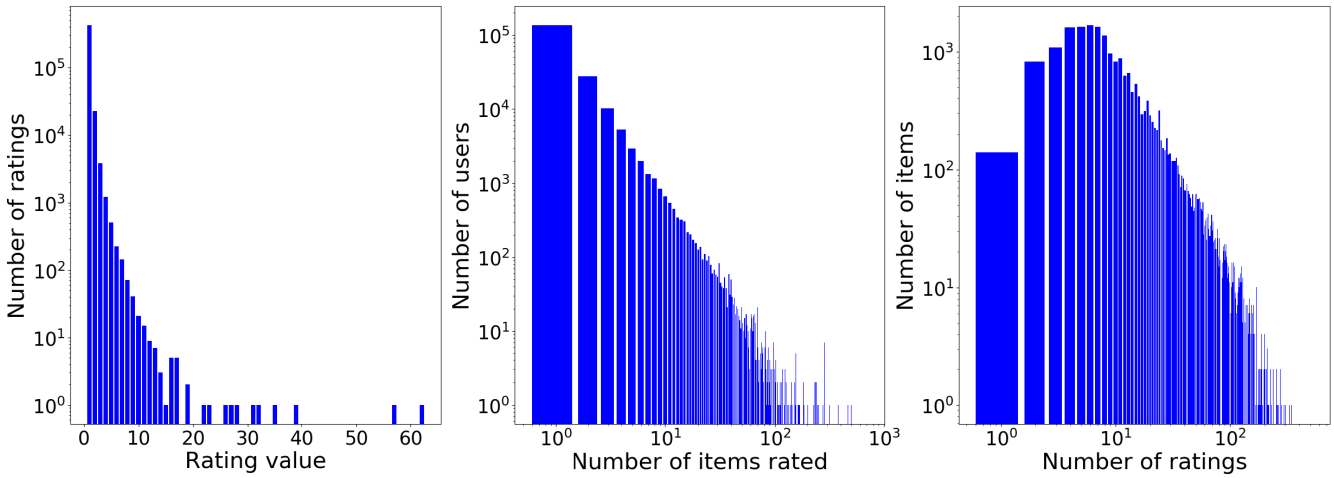
The total number of items rated by user is represented on the center graphics of Figures 5a and 5b. For example, for ARM, 5207 authors have only one item rated (cold start problem), and for CheRM this value is 136,391 authors. In our context, this means that 30% of the authors in ARM only wrote about one of the cluster of stars of our list and 70% on the authors in CheRM only wrote about one Chemical Compound of our list. The right graphics of Figures 5a and 5b show the number of ratings by item. There are no items with only one rating for ARM, with the minimal number of

TABLE 3. Dimensions of the datasets evaluated in this study for nUsers (number of users), nItems (number of items), nRat (number of ratings), minRat (minimal rating), maxRat (maximum rating), nThreshold (number of ratings greater or equal to the defined threshold), sparsity, mean, SD (Standard deviation), mode, and median.

Dataset	nUsers	nItems	nRatings	minRat	maxRat	nThreshold	Sparsity	Mean	SD	Mode	Median
cheRM	193106	22299	456681	1	62	28967	99.98	1.08	0.45	1	1
cheRM-20	2193	16437	117020	1	62	7600	99.67	1.08	0.41	1	1
ARM	17006	2102	179269	1	89	49325	99.49	1.72	2.02	1	1
ARM-20	1493	2101	106104	1	89	34258	96.61	1.92	2.37	1	1
ML-100k	943	1682	100000	1	5	21201	93.69	3.52	1.12	4	4
SD4AI	14143	18502	1389094	1	160	216746	99.46	2.38	2.61	1	1.75



(a) Analysis of ARM dataset. Left: Distribution of rating values; Center: Number of rated items by user; Right: Number of ratings by item.



(b) Analysis of CheRM dataset. Left: Distribution of rating values; Center: Number of rated items by user; Right: Number of ratings by item.

FIGURE 5. Visual analysis of the datasets ARM and CheRM.

ratings being 2, for a total of 6 items. There are 176 items with 11 ratings each, and this is the most frequent number of ratings. The most rated item is "Melotte 22", with ratings from 5287 users. For CheRM, there are 140 Chemical Compounds with only one rating, and the item with more ratings is CHEBI:465284 (ganciclovir) with ratings from 529 authors.

B. DATASET VALIDATION

To elucidate about what is being recommended with ARM, Figure 6 provides an example of what the PIP algorithm recommends to user 1206 in a top 10 ranked list. Thus, for this user using PIP, the ranked list of recommended items is [175, 187, 1104, 1139, 1850, 152, 1573, 2002, 2012 and 866], which corresponds to the OCs named [Melotte 20, IC 348, IC 2602, NGC 3532, Roslund 5, NGC 1039, NGC 6494, NGC 7092, Trumpler 37, and NGC 2571], respectively. The OCs underlined are the ones correctly recommended, i.e., relevant for this user (according to the previously defined threshold of 2.0). For this user, the Precision is 0.60, and the Recall is 0.86 since this user has 7 relevant items in the test set. For CheRM, instead of OCs, we are recommending Chemical Compounds from ChEBI.

TABLE 4. Recommender algorithms top results for each evaluation metric, for ARM, SD4AI, ARM-20, CheRM-20 and ML-100k datasets.

	ARM		SD4AI		ARM-20		CheRM-20		ML-100k	
M	**Algo**	**Value**	**Algo**	**Value**	**Algo**	**Value**	**Algo**	**Value**	**Algo**	**Value**
MAE	JMSD	0.593	COR	0.562	JMSD	0.903	MSD	0.100	PIP	0.754
PRE	COR	0.371	PIP	0.641	JAC	0.599	PIP	0.158	JAC	0.356
REC	PIP	0.936	PIP	0.793	PIP	0.893	PIP	0.609	PIP	0.705
F1	COS	0.770	PIP	0.601	JMSD	0.600	COS	0.740	PIP	0.425
nDCG	PIP	0.845	PIP	0.769	PIP	0.838	PIP	0.836	PIP	0.773

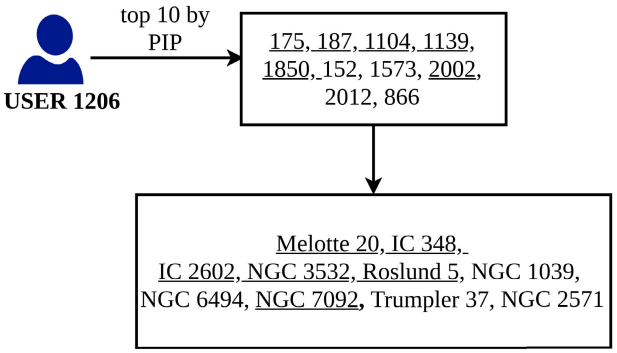


FIGURE 6. Example of the top 10 recommendations of open clusters to the user 1206, calculated by the PIP recommender algorithm. The OCs underlined are the ones from the top 10 that are relevant to the user 1206.

Table 4 shows the top results obtained for ARM, ARM-20 and CheRM-20, as well as for SD4AI and ML-100k, for the measures (**M**) MEA, PRE, REC, F1, and nDCG. The table presents the maximum value for each measure (**Value**), and the algorithm where it was obtained (**Algo**). The algorithms in question are COR, COS, PIP, JMSD, JAC, and MSD. The results of PMF are presented separately for a better comparison of Memory-based and Model-based algorithms. It was not possible to get the results for the full CheRM dataset since CF4J cannot process datasets of such large dimensions, thus we used CheRM-20 for a fairer comparison with ML-100k and ARM-20. ARM achieved better results for Recall, F-measure and nDCG than SD4AI. For ARM-20 the results of the Precision are better than for ARM. Due to its dimensions (Table 3), ARM-20 is more comparable to ML-100k, and its results for Precision, Recall, F-measure, and nDCG are higher. PIP is the recommender algorithm that achieved the best results for most of the evaluation measures in all datasets. The Precision for ARM is the value that presents a higher difference for the highest Precision achieved with SD4AI. MAE is similar in all datasets, however this measure is not directly comparable due to the different range of ratings values of the evaluation datasets.

Figures 7, 8 and 9 show in more detail the results of Precision, Recall, and nDCG, respectively, obtained in the different datasets with the different algorithms. Analysing the plots, we see that the datasets present similar behavior for the same algorithms.

PMF results are presented in Table 5. ARM-20 benefits from this algorithm only for Recall and nDCG. For ML-100k and SD4AI, PMF is the recommender algorithm with the best results. Thus, based on these results, we can say that using CF4J, Memory-based algorithms work better than Model-based algorithms for the ARM dataset.

TABLE 5. Results for the PMF recommender algorithm for the datasets ARM-20, CheRM-20, SD4AI, and ML-100k.

Measure	ARM-20	CheRM-20	ML-100k	SD4AI
MAE	0.906	0.200	0.756	0.686
PRE	0.501	0.149	0.479	0.669
REC	0.923	0.893	0.774	0.886
F1	0.569	0.726	0.466	0.609
nDCG	0.886	0.955	0.832	0.815

The results for each dataset are not directly comparable since they use similar but not equal settings (e.g.: minimum and maximum rating, thresholds), however they provide sound indication of LIBRETTI effectiveness.

V. DISCUSSION

The lack of datasets for deploying or evaluating recommender algorithms for scientific data exploration is a major drawback delaying their use and development in this area. The proposed methodology, LIBRETTI, is a solution for the lack of ratings, taking advantage of the comprehensive list of scientific publications available for all research areas.

From the results presented in Section IV-A we can say that ARM, ARM-20 and CheRM-20 are similar to other datasets often used in the field of RSs, such as movies datasets, with similar values of data sparsity. Compared with SD4AI, ARM has less items, however, we achieved almost the same number of users. This is a positive point because we will have more users to search for similarity. A disadvantage of ARM is its high percentage of users who rated only a few items. For instance, when we remove the users who rated less than 20 items to create ARM-20, the number of users is reduced to less than 8%. In the case of CheRM-20, the number of users is reduced to 1.13% of the original dataset. This may be mitigated by using NER to extract more items from each article, items that may not be identified in the external sources of knowledge that we used (SIMBAD, ADS and ChEBI). Like this, we will have more items rated for the same number of users. Despite that, the results with ARM-20 are strong, as may be seen in Table 4 and 5. For Precision, Recall, F-measure and nDCG, ARM-20 results are higher than the results of ML-100k. ML-100k is a dataset widely used for

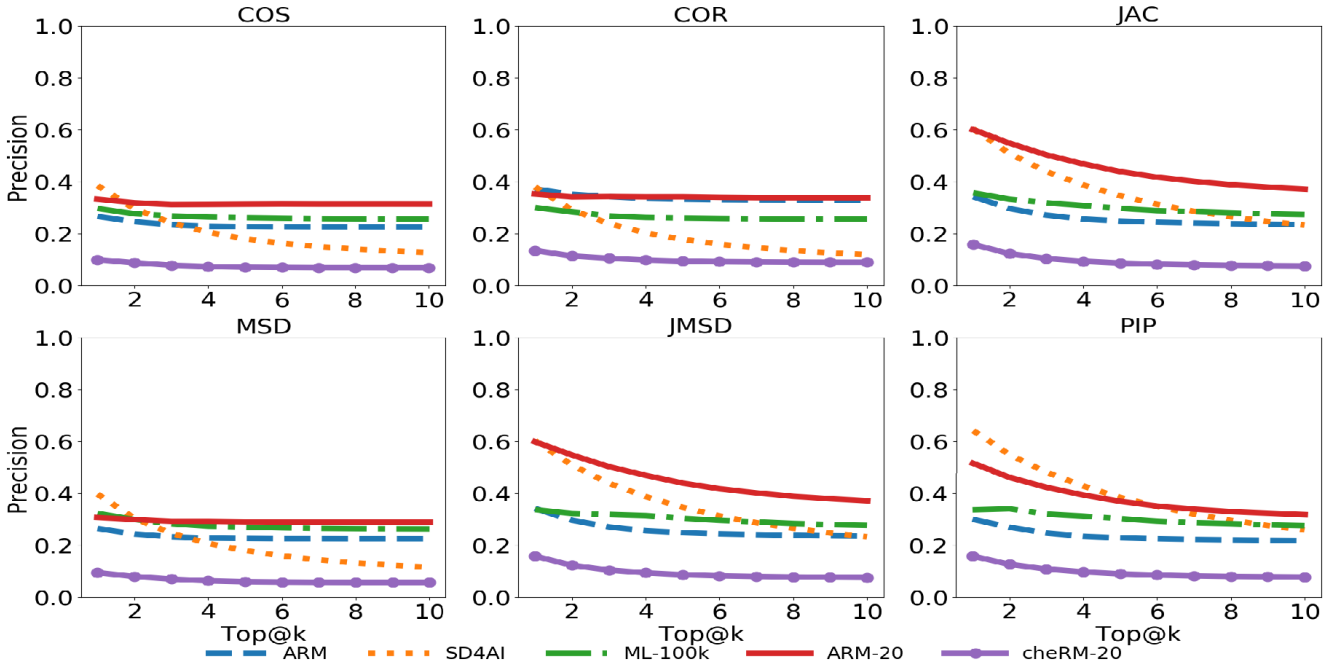


FIGURE 7. Precision at k (k between 1 and 10), for pearson correlation, cosine similarity, jaccard index, jaccard mean squared difference, mean squared difference, and proximity impact popularity, for ARM, ARM-20, CheRM, ML-100k and SD4AI datasets.

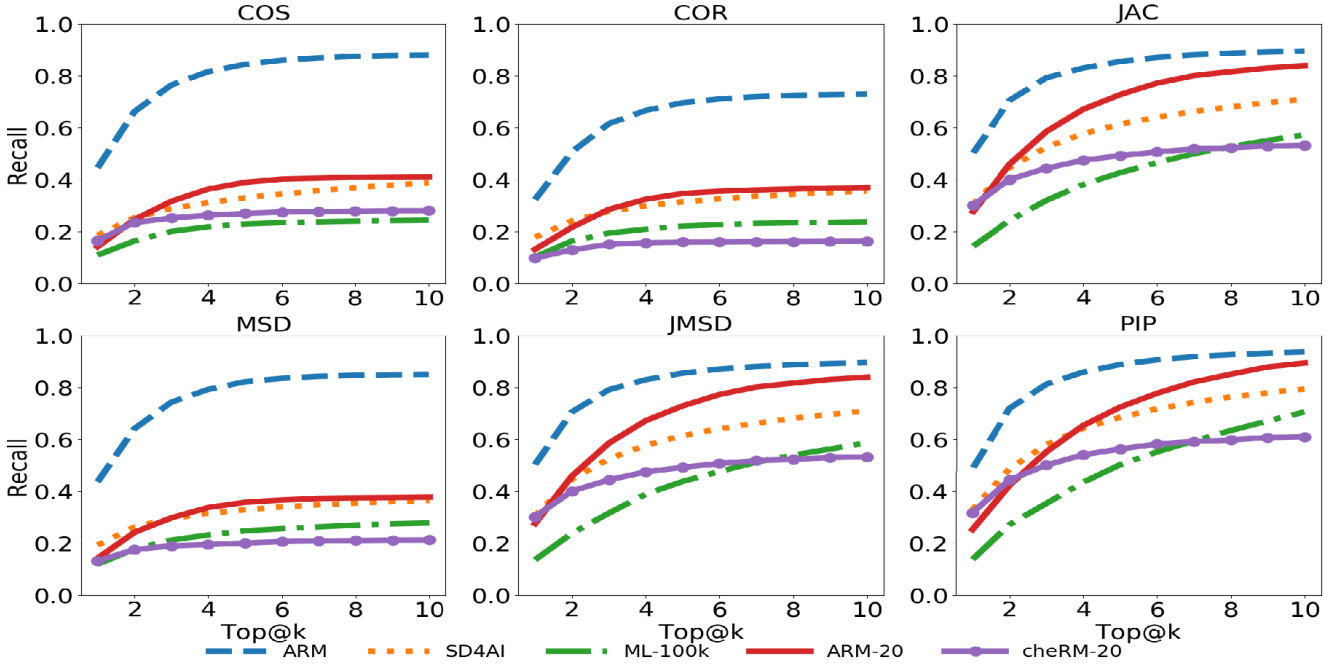


FIGURE 8. Recall at k (k between 1 and 10), for pearson correlation, cosine similarity, jaccard index, jaccard mean squared difference, mean squared difference, and proximity impact popularity, for ARM, ARM-20, CheRM, ML-100k and SD4AI datasets.

evaluating recommender algorithms, thus these results support our hypothesis that ARM is a viable solution in assessing recommender algorithms in scientific fields.

The results for CheRM-20, particularly for precision, are lower than the results for the other datasets in this study. This may be explained by the fact that CF4J is a framework more suitable for datasets of explicit data, where we can define a threshold for the rating, defining an item as relevant/not relevant. The datasets developed through LIBRETTI methodology are implicit and even the minimal rating, 1, is relevant.

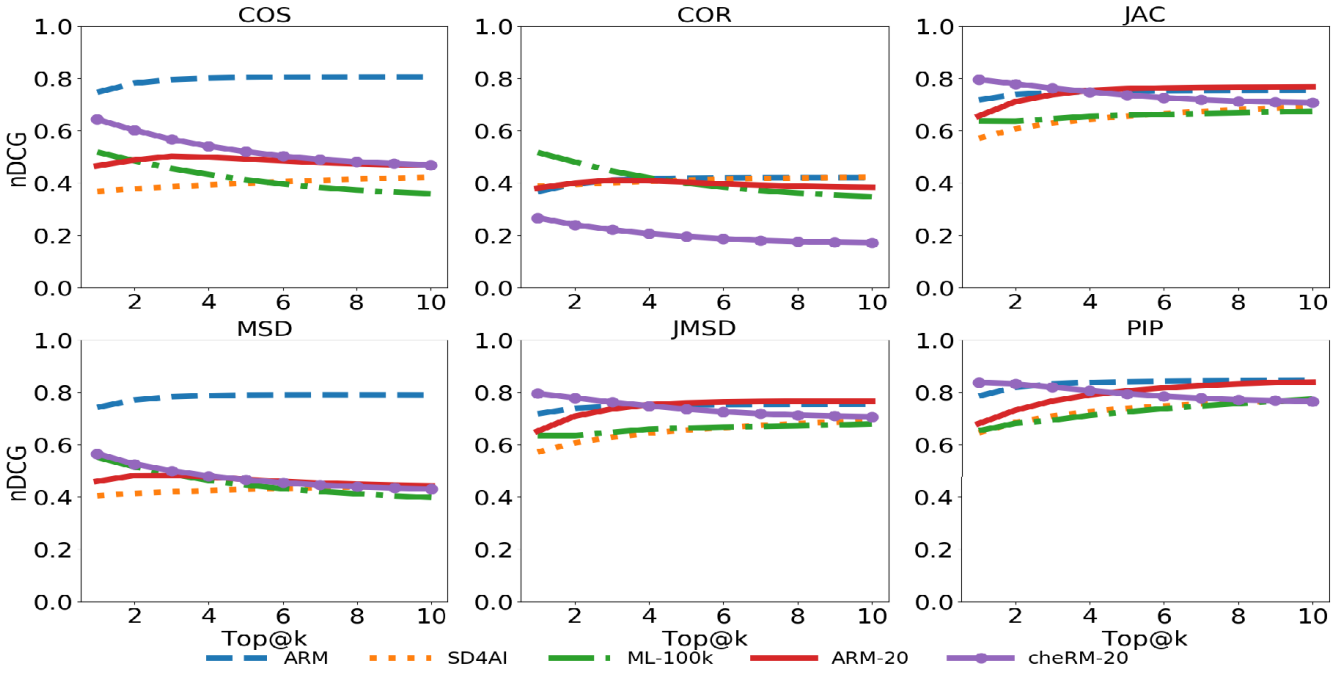


FIGURE 9. nDCG at k (k between 1 and 10), for pearson correlation, cosine similarity, jaccard index, jaccard mean squared difference, mean squared difference, and proximity impact popularity, for ARM, ARM-20, CheRM, ML-100k and SD4AI datasets.

When we move the threshold to 2, we are losing, in the case of CheRM, 93% of the actually relevant ratings (see Table 3). Thus, for example, if the RS recommends 5 items, whose real ratings are 1, the Precision will be zero, since all the ratings are bellow the defined threshold. If we define the threshold as 1, the Precision will always be one, since CF4J only recommends items from the testset that we have a real rating. For example, if a user in the testset rated 10 items, and we want the top 5, CF4J ranks only these 10 items and recommends 5 of them. As the threshold is 1, the Precision is 1 because all items are relevant.

An advantage of the datasets created with LIBRETTI is that they may be used as direct input data for CF platforms, mitigating the sparsity problem. The pure cold start problem of new users, which do not have any rated item, is not overcome by our datasets. However, with a few ratings we can easily find similar users. The cold start for new items is also a challenge in CF. Our dataset may help solving this problem by introducing into the recommendation platforms implicit ratings for these unrated items.

The datasets created with LIBRETTI can also be used for testing and evaluating recommender algorithms. The dataset is filled with real people (the authors of the articles), who in one moment of their research had interest for that item they mentioned. For example, if we were evaluating which is the best algorithm for recommending OCs, analysing Figures 7, 8 and 9, for Precision it would be JMSD, and for Recall and nDCG it would be PIP. Another advantage is that LIBRETTI is scalable, and not limited to a small number of items. The most limiting point related to the scalability of the method is the access restrictions that may be imposed by the external sources. For example, the ADS API only allows 5000 requests per day. Another advantage of LIBRETTI is that the database creation process runs offline. Thus, it does not interfere with the retrieval of the recommendation to the user, and it is easy to keep updated, with regular crawling for new articles.

The application of LIBRETTI for creating ARM and CheRM is fully available at https://github.com/lasigeBioTM/cARM and https://github.com/lasigeBioTM/CheRM, as well as the datasets used in this study.

VI. CONCLUSION

The main goal of this work was to provide a validated methodology for generating datasets of implicit feedback suitable for recommending scientific items using CF approaches. The proposed methodology, LIBRETTI, consists in identifying a list of items/objects, finding research articles mentioning each item, extracting the authors from each article, and finally creating a <user,item,rating> dataset where users are unique authors from the collected articles, and the rating values are the number of articles a unique author wrote about an item. We used Astronomy and Chemistry as case studies and compared the obtained datasets (ARM, ARM-20 and CheRM-20) with SD4AI and ML-100k. Considering the results obtained, we believe that LIBRETTI paves the way to a widely applicable and an effective solution for testing and evaluating the use of recommender algorithms

in scientific areas and for the recommendation of not studied items for the researchers.

In future work, we shall further explore the performance of LIBRETTI with algorithms designed for implicit datasets, as well as its use in CB and Hybrid algorithms using the features of the items.

ACKNOWLEDGMENT

This research has made use of NASA's Astrophysics Data System Bibliographic Services, SIMBAD database, operated at CDS, Strasbourg, France, Chemical Entities of Biological Interest (ChEBI), and PubMed.

REFERENCES

[1] L. Bornmann and R. Mutz, "Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references," *J. Assoc. Inf. Sci. Technol.*, vol. 66, no. 11, pp. 2215–2222, 2015.

[2] F. Ricci, L. Rokach, and B. Shapira, "Recommender systems: Introduction and challenges," in *Recommender Systems Handbook*. Boston, MA, USA: Springer, 2015, pp. 1–34.

[3] S. Sivapalan, A. Sadeghian, H. Rahnama, and A. M. Madni, "Recommender systems in e-commerce," in *Proc. IEEE World Automat. Congr. (WAC)*, Aug. 2014, pp. 179–184. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6935763/authors#authors

[4] J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang, "Recommender system application developments: A survey," *Decis. Support Syst.*, vol. 74, pp. 12–32, Jun. 2015.

[5] J. Beel, B. Gipp, S. Langer, and C. Breitinger, "Paper recommender systems: A literature survey," *Int. J. Digit. Libraries*, vol. 17, no. 4, pp. 305–338, Nov. 2016.

[6] R. Burke, "Hybrid recommender systems: Survey and experiments," *User Model. User-Adapted Interact.*, vol. 12, no. 4, pp. 331–370, Nov. 2002.

[7] S. Perugini, M. A. Gonçalves, and E. A. Fox, "Recommender systems research: A connection-centric survey," *J. Intell. Inf. Syst.*, vol. 23, no. 2, pp. 107–143, 2004.

[8] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, Jun. 2005.

[9] N. Tintarev and J. Masthoff, "A survey of explanations in recommender systems," in *Proc. IEEE 23rd Int. Conf. Data Eng. Workshop*, Apr. 2007, pp. 801–810.

[10] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Adv. Artif. Intell.*, vol. 2009, pp. 4:2–4:2, Aug. 2009.

[11] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowl. Syst.*, vol. 46, pp. 109–132, Jul. 2013, doi: 10.1016/j.knosys.2013.03.012.

[12] I. Portugal, P. Alencar, and D. Cowan, "The use of machine learning algorithms in recommender systems: A systematic review," *Expert Syst. Appl.*, vol. 97, pp. 205–227, May 2018.

[13] Z. Batmaz, A. Yurekli, A. Bilge, and C. Kaleli, "A review on deep learning for recommender systems: Challenges and remedies," *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 1–37, Jun. 2018.

[14] F. Alyari and N. J. Navimipour, "Recommender systems: A systematic review of the state of the art literature and suggestions for future research," *Kybernetes*, vol. 47, no. 5, pp. 985–1017, 2018.

[15] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.

[16] G. Shani and A. Gunawardana, "Evaluating recommendation systems," in *Recommender Systems Handbook*. F. Ricci, L. Rokach, B. Shapira, and P. Kantor, Eds. Boston, MA, USA: Springer, 2011.

[17] A. C. Valdez, M. Ziefle, K. Verbert, A. Felfernig, and A. Holzinger, "Recommender systems for health informatics: State-of-the-art and future perspectives," in *Machine Learning for Health Informatics*. Cham, Switzerland: Springer, 2016, pp. 391–414.

[18] H. Schäfer, S. Hors-Fraile, R. P. Karumur, A. Calero Valdez, A. Said, H. Torkamaan, T. Ulmer, and C. Trattner, "Towards health (aware) recommender systems," in *Proc. Int. Conf. Digit. health*, 2017, pp. 157–161.

[19] W. Zhang, H. Zou, L. Luo, Q. Liu, W. Wu, and W. Xiao, "Predicting potential side effects of drugs by recommender methods and ensemble learning," *Neurocomputing*, vol. 173, pp. 979–987, Jan. 2016.

[20] A. Wang, H. Lim, S.-Y. Cheng, and L. Xie, "ANTENNA, a multi-rank, multi-layered recommender system for inferring reliable drug-gene-disease associations: Repurposing diazoxide as a targeted anti-cancer therapy," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 15, no. 6, pp. 1960–1967, Nov./Dec. 2018.

[21] C. Suphavilai, D. Bertrand, and N. Nagarajan, "Predicting cancer drug response using a recommender system," *Bioinformatics*, vol. 34, no. 22, pp. 3907–3914, 2018.

[22] H. C. Wittich, M. Seeland, and J. Wäldchen, M. Rzanny, and P. Mäder, "Recommending plant taxa for supporting on-site species identification," *BMC bioinformatics*, vol. 19, no. 1, p. 190, 2018.

[23] M. Ge, F. Ricci, and D. Massimo, "Health-aware food recommender system," in *Proc. 9th ACM Conf. Recommender Syst.*, 2015, pp. 333–334.

[24] D. Jannach, M. Zanker, M. Ge, and M. Gröning, "Recommender systems in computer science and information systems—A landscape of research," in *E-Commerce and Web Technologies* (Lecture Notes in Business Information Processing), vol. 123, C. Huemer and P. Lops, Eds. Berlin, Germany: Springer, 2012.

[25] F. Ortega, J. Bobadilla, A. Gutiérrez, R. Hurtado, and X. Li, "Artificial intelligence scientific documentation dataset for recommender systems," *IEEE Access*, vol. 6, pp. 48543–48555, 2018.

[26] W. Dias, B. Alessi, A. Moitinho, and J. Lépine, "New catalogue of optically visible open clusters and candidates," *Astron. Astrophys.*, vol. 389, no. 3, pp. 871–873, 2002.

[27] M. Wenger, F. Ochsenbein, D. Egret, P. Dubois, F. Bonnarel, S. Borde, F. Genova, G. Jasniewicz, S. Laloë, S. Lesteven, and R. Monier, "The SIMBAD astronomical database: The CDS reference database for astronomical objects," *Astron. Astrophys. Suppl. Ser.*, vol. 143, no. 1, pp. 9–22, 2000.

[28] M. J. Kurtz, G. Eichhorn, A. Accomazzi, C. Grant, M. Demleitner, and S. S. Murray, "Worldwide use and impact of the nasa astrophysics data system digital library," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 56, no. 1, pp. 36–45, 2005.

[29] K. Degtyarenko, P. De Matos, M. Ennis, J. Hastings, M. Zbinden, A. McNaught, R. Alcántara, M. Darsow, M. Guedj, and M. Ashburner, "ChEBI: A database and ontology for chemical entities of biological interest," *Nucleic Acids Res.*, vol. 36, suppl. 1, pp. D344–D350, Jan. 2008, doi: 10.1093/nar/gkm791.

[30] F. Ortega, B. Zhu, J. Bobadilla, and A. Hernando, "CF4J: Collaborative filtering for Java," *Knowl.-Based Syst.*, vol. 152, pp. 94–99, Jul. 2018.

[31] S. Zahra, M. A. Ghazanfar, A. Khalid, M. A. Azam, U. Naeem, and A. Prugel-Bennett, "Novel centroid selection approaches for KMeans-clustering based recommender systems," *Inf. Sci.*, vol. 320, pp. 156–189, Nov. 2015.

[32] G. Guo, J. Zhang, and N. Yorke-Smith, "Leveraging multiviews of trust and similarity to enhance clustering-based recommender systems," *Knowl.-Based Syst.*, vol. 74, pp. 14–27, Jan. 2015.

[33] A. L. V. Pereira and E. R. Hruschka, "Simultaneous co-clustering and learning to address the cold start problem in recommender systems," *Knowl.-Based Syst.*, vol. 82, pp. 11–19, Jul. 2015.

[34] A. Hernando, J. Bobadilla, and F. Ortega, "A non negative matrix factorization for collaborative filtering recommender systems based on a Bayesian probabilistic model," *Knowl.-Based Syst.*, vol. 97, pp. 188–202, Apr. 2016.

[35] F. Alqadah, C. K. Reddy, J. Hu, and H. F. Alqadah, "Biclustering neighborhood-based collaborative filtering method for top-n recommender systems," *Knowl. Inf. Syst.*, vol. 44, no. 2, pp. 475–491, 2015.

[36] F. Zhang, N. J. Yuan, D. Lian, X. Xie, and W.-Y. Ma, "Collaborative knowledge base embedding for recommender systems," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 353–362.

[37] T. Ishihara, Y. Koga, Y. Iwatsuki, and F. Hirayama, "Identification of potent orally active factor Xa inhibitors based on conjugation strategy and application of predictable fragment recommender system," *Bioorganic Medicinal Chem.*, vol. 23, no. 2, pp. 277–289, 2015.

[38] A. Seko, H. Hayashi, and I. Tanaka, "Compositional descriptor-based recommender system for the materials discovery," *J. Chem. Phys.*, vol. 148, no. 24, p. 241719, 2018.

[39] A. B. Owen, J. Stuart, K. Mach, A. M. Villeneuve, and S. Kim, "A gene recommender algorithm to identify coexpressed genes in C. Elegans," *Genome Res.*, vol. 13, no. 8, pp. 1828–1837, 2003.

[40] D. M. Ng, M. H. Woehrmann, and J. M. Stuart, "Recommending pathway genes using a compendium of clustering solutions," in *Biocomputing*. Singapore: World Scientific, 2007, pp. 379–390.

[41] A. Torkaman, N. M. Charkari, and M. Aghaeipour, "An approach for leukemia classification based on cooperative game theory," *Anal. Cellular Pathol.*, vol. 34, no. 5, pp. 235–246, 2011.

[42] J. Bostrøm, N. Falk, and C. Tyrchan, "Exploiting personalized information for reagent selection in drug design," *Drug Discovery Today*, vol. 16, nos. 5–6, pp. 181–187, 2011.
[43] W. De Smet, K. De Loof, P. De Vos, P. Dawyndt, and B. De Baets, "Filtering and ranking techniques for automated selection of high-quality 16S rRNA gene sequences," *Systematic Appl. Microbiol.*, vol. 36, no. 8, pp. 549–559, 2013.
[44] M. Wiesner and D. Pfeifer, "Health recommender systems: Concepts, requirements, technical basics and challenges," *Int. J. Environ. Res. Public Health*, vol. 11, no. 3, pp. 2580–2607, 2014.
[45] J. H. Chen, T. Podchiyska, and R. B. Altman, "OrderRex: Clinical order decision support and outcome predictions by data-mining electronic medical records," *J. Amer. Med. Inform. Assoc.*, vol. 23, no. 2, pp. 339–348, 2015.
[46] A. A. Macedo, J. T. Pollettini, J. A. Baranauskas, and J. C. A. Chaves, "A health surveillance software framework to deliver information on preventive healthcare strategies," *J. Biomed. Inform.*, vol. 62, pp. 159–170, 2016.
[47] F. Hao and R. H. Blair, "A comparative study: Classification vs. user-based collaborative filtering for clinical prediction," *BMC Med. Res. Methodol.*, vol. 16, no. 1, p. 172, 2016.
[48] G. Corrado, T. Tebaldi, F. Costa, P. Frasconi, and A. Passerini, "RNAcommender: Genome-wide recommendation of rna–protein interactions," *Bioinformatics*, vol. 32, no. 23, pp. 3627–3634, 2016.
[49] A. Mustaqeem, S. M. Anwar, A. R. Khan, and M. Majid, "A statistical analysis based recommender model for heart disease patients," *Int. J. Med. Inform.*, vol. 108, pp. 134–145, Dec. 2017.
[50] C. L. S. Bocanegra, J. L. S. Ramos, C. Rizo, A. Civit, and L. Fernandez-Luque, "Healthrecsys: A semantic content-based recommender system to complement health videos," *BMC Med. Inform. Decis. Making*, vol. 17, no. 1, p. 63, 2017.
[51] F. Gräßer, S. Beckert, D. Küster, J. Schmitt, S. Abraham, H. Malberg, and S. Zaunseder, "Therapy decision support based on recommender system methods," *J. Healthcare Eng.*, vol. 2017, Mar. 2017, Art. no. 8659460, doi: 10.1155/2017/8659460.
[52] J. H. Chen, M. Alagappan, M. K. Goldstein, S. M. Asch, and R. B. Altman, "Decaying relevance of clinical data towards future decisions in data-driven inpatient clinical order sets," *Int. J. Med. Inf.*, vol. 102, pp. 71–79, Jun. 2017.
[53] L. Yang, C.-K. Hsieh, H. Yang, J. P. Pollak, N. Dell, S. Belongie, C. Cole, and D. Estrin, "Yum-Me: A personalized nutrient-based meal recommender system," *ACM Trans. Inf. Syst.*, vol. 36, no. 1, p. 7, 2017.
[54] E. Pustozerov, P. Popova, A. Tkachuk, Y. Bolotko, Z. Yuldashev, and E. Grineva, "Development and evaluation of a mobile personalized blood glucose prediction system for patients with gestational diabetes mellitus," *JMIR mHealth uHealth*, vol. 6, no. 1, p. e6, 2018.
[55] G. Agapito, M. Simeoni, B. Calabrese, and I. Caré, T. Lamprinoudi, P. H. Guzzi, A. Pujia, G. Fuiano, and M. Cannataro, "DIETOS: A dietary recommender system for chronic diseases monitoring and management," *Comput. Methods Programs Biomed.*, vol. 153, pp. 93–104, Jan. 2018.
[56] J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger, "DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network," *BMC Med. Res. Methodol.*, vol. 18, no. 1, p. 24, 2018.
[57] F. Torrent-Fontbona and B. López, "Personalized adaptive CBR bolus recommender system for type 1 diabetes," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 1, pp. 387–394, Jan. 2019.
[58] A. Accomazzi, M. J. Kurtz, E. A. Henneken, C. S. Grant, D. M. Thompson, R. Chyla, S. McDonald, T. J. Shaulis, S. Blanco-Cuaresma, and G. Shapurian, "New ADS functionality for the curator," in *Proc. Eur. Phys. J. Web Conf.*, vol. 186, Jul. 2018, Art. no. 08001.
[59] W. E. Kerzendorf, "Knowledge discovery through text-based similarity searches for astronomy literature," *J. Astrophys. Astron.*, vol. 40, no. 3, p. 23, 2019.
[60] N. Mukund, S. Thakur, S. Abraham, A. Aniyan, S. Mitra, N. S. Philip, K. Vaghmare, and D. Acharjya, "An information retrieval and recommendation system for astronomical observatories," *Astrophys. J. Suppl. Ser.*, vol. 235, no. 1, p. 22, 2018.
[61] F. M. Harper and J. A. Konstan, "The movielens datasets: History and context," *ACM Trans. Interact. Intell. Syst.*, vol. 5, no. 4, p. 19, Jan. 2016.
[62] M. D. Ekstrand, "The LKPY package for recommender systems experiments: Next-generation tools and lessons learned from the LensKit project," Sep. 2018, *arXiv:1809.03125*. [Online]. Available: https://arxiv.org/abs/1809.03125
[63] D. Lyubimov and A. Palumbo, *Apache Mahout: Beyond MapReduce*. Scotts Valley, CA, USA: CreateSpace Independent Publishing Platform, 2016.
[64] J. Bobadilla, F. Serradilla, and J. Bernal, "A new collaborative filtering metric that improves the behavior of recommender systems," *Knowl.-Based Syst.*, vol. 23, no. 6, pp. 520–528, 2010.
[65] H. Wang, N. Wang, and D.-Y. Yeung, "Collaborative deep learning for recommender systems," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2015, pp. 1235–1244.
[66] U. Panniello, A. Tuzhilin, and M. Gorgoglione, "Comparing context-aware recommender systems in terms of accuracy and diversity," *User Model. User-Adapted Interact.*, vol. 24, nos. 1–2, pp. 35–65, 2014.

MÁRCIA BARROS was born in 1990. She received the B.S. degree in biomedical sciences from the Universidade da Beira Interior, Covilhã, Portugal, in 2012, the M.Sc. degree in legal medicine and forensic sciences from the Universidade de Coimbra, Portugal, in 2014. She is currently pursuing the Ph.D. degree in informatics with the Universidade de Lisboa, Portugal. She has worked with astronomical data, mainly with data from the European Space Agency's Gaia mission, and contributed to the creation of high-quality images of the Milky Way. Her current topics of research are recommender systems for scientific data, machine learning, and text mining.

ANDRÉ MOITINHO received the B.Sc. and M.Sc. degrees in physics from the University of Lisbon, Portugal, and the Ph.D. degree in astrophysics from the University of Granada, Spain, in 1999. He is currently an Assistant Professor with the Universidade de Lisboa, where he leads a research group in astronomy and associated technologies. His research interests include big data visualization, instrumentation for astronomy, galactic astronomy, star formation, and long term sustainability of space activities. He leads the development of the visualization services for the Gaia mission of the European Space Agency (ESA). He has served several committees, including the ESA Astronomy Working Group, and the Scientific and Technical Committee of the European Sothern Observatory (ESO). He is currently the Chairperson of the Board of Directors of the journal *Astronomy & Astrophysics* (Impact Factor 6.209).

FRANCISCO M. COUTO received the degree and the master's degree in informatics and computer engineering from IST, in 2000 and 2001, respectively, and the Ph.D. degree in informatics, specialization bioinformatics, from the Universidade de Lisboa, in 2006. He was with the faculty at IST, from 1998 to 2001. Since 2001, he has been with the Faculdade de Ciências, Universidade de Lisboa (FCUL). He was an Invited Researcher with EBI, AFMB-CNRS, and BioAlma during his Ph.D. studies. He is currently an Associate Professor with habilitation and the Head of the Department of Informatics, FCUL, a member of coordination board of the master in bioinformatics and computational biology, and a member of LASIGE, where he is the Coordinator and a Founder of the HBI: Health and Biomedical Informatics research line.

He received the Young Engineer Innovation Prize 2004 from the Portuguese Engineers Guild, and an Honorable Mention in 2017 and the prize in 2018 of the ULisboa/Caixa Geral de Depósitos (CGD) Scientific Prizes.