IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 29, 2018, accepted December 16, 2018, date of publication December 28, 2018, date of current version January 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2890079

A New Recommendation Approach Based on Probabilistic Soft Clustering Methods: A Scientific Documentation Case Study

REMIGIO HURTADO[1,2], JESÚS BOBADILLA[2], RODOLFO BOJORQUE[1,2], FERNANDO ORTEGA[2], AND XIN LI[3]

[1]Universidad Politécnica Salesiana, Sistemas y Computación, Cuenca 010102, Ecuador
[2]Universidad Politécnica de Madrid, Intelligent Systems, 28031 Madrid, Spain
[3]Beijing Institute of Technology, School of Computer Science, Beijing 100081, China

Corresponding author: Remigio Hurtado (rhurtadoo@ups.edu.ec)

ABSTRACT Recommender system (RS) clustering is an important issue, both for the improvement of the collaborative filtering (CF) accuracy and to obtain analytical information from their high sparse datasets. RS items and users usually share features belonging to different clusters, e.g., a musical-comedy movie. Soft clustering, therefore, is the CF clustering's most natural approach. In this paper, we propose a new prediction approach for probabilistic soft clustering methods. In addition, we put to test a not traditional scientific documentation CF dataset: SD4AI, and we compare results with the MovieLens baseline. Not traditional CF datasets have challenging features, such as not regular rating frequency distributions, broad range of rating values, and a particularly high sparsity. The results show the suitability of using soft-clustering approaches, where their probabilistic overlapping parameters find optimum values when balanced hard/soft clustering is used. This paper opens some promising lines of research, such as RSs' use in the scientific documentation field, the Internet of Things-based datasets processing, and design of new model-based soft clustering methods.

INDEX TERMS Soft clustering, scientific documentation, collaborative filtering, recommender systems.

I. INTRODUCTION

Recommender Systems (RS) [1], [2] are part of the artificial intelligence current set of tools to face the information overload problem. RS users can cast explicit ratings showing their preferences related to an extensive set of items, such as: movies [3], music [4], books [5], etc. There are also RS that record user's preferences based on their interaction with the System (mouse clicks, visited items, consumed items, etc.). RS can be mainly classified as content, demographic, social, context aware and collaborative. Commercial RS usually implement a hybrid approach of several of the mentioned types. Modern RS are based on the Collaborative Filtering (CF) [1], [6] approach.

CF RS make recommendation to users based on their preferences from the complete set of ratings. The traditional approach was the memory-based one; it is usually implemented using the KNN algorithm [7]. Memory-based methods make recommendations directly based on ratings. Current CF approaches are model-based [8]: first a model is created from the data; and then, recommendations are made from the model. Matrix Factorization (MF) methods [9]–[11] are the most used model-based ones; they are simple, efficient and accurate.

There is a big amount of publications referring to the prediction and recommendation processes in CF RS; These papers make use of traditional datasets, usually: Movielens, Filmtrust, Jester, Netflix, etc. There is also a variety of publications that are based on the hard clustering of CF RS, especially carrying out the recommendation process after the clustering phase [12], [13]. There are several publications of soft clustering methods using the Fuzzy C Means algorithm (FCM). Bezdek *et al.* [14] provide a Fuzzy C-Means implementation. It aggregates subsets using a generalized least-squares objective function. Bezdek [15] focuses on the Fuzzy C-Means clustering applied to the pattern recognition field. The convergence of these types of algorithms is also deeply described. Wu and Yang [16] show a limitation of FCM in RS with high dimensionality datasets. Zahra *et al.* [12] improve the quality of FCM with respect to the traditional FCM by means of initialization variants of centroids. Its prediction approach is that the user-item ratings matrix is clustered in K clusters, then predictions and recommendations are provided based on the ratings of similar centroids. The similarity between the active user and all K centroids is calculated using Pearson correlation coefficient. A better recommendation results compared to other clustering methods is achivied in [17] by combining Center of Gravity defuzzified Fuzzy Clustering and Pearson correlation coefficient get better recommendation results compared to other clustering methods. Its approach executes the recommendation analyzing only the group to which the user belongs, this means that the predictions and recommendations are provided based on the ratings of similar users of the cluster to which the active user belongs using Pearson correlation coefficient as a measure of similarity. They show that the best accuracy and precision results are achieved with a low K configuration (3 clusters), while that performing greater cluster number will lead to lower accuracy and precision. A low K does not provide enough knowledge to perform data analytics. References [12], [14], [15], and [17] have a promising route to improve the prediction quality results.

Therefore, existing research in CF RS clustering focuses on the use of traditional datasets and, hard and soft clustering methods. However, in this paper an innovative research is carried out in three aspects:

1) Design of a new prediction approach for probabilistic soft clustering methods.
2) Use of a not traditional dataset: SD4AI [18], which contains information on scientific documentation. The most relevant differences between this dataset and the traditional ones are: a) A much wider range of ratings; 640 possible values, as opposed to the usual ones: 1, 2, 3, 4, 5 or like, dislike, b) A non-regular frequency distribution of the ratings, c) A higher level of sparsity than traditional datasets of its size.
3) Use of soft-clustering techniques, which allow to establish different degrees of overlapping (probabilistic assignment of each element to several classes).

The main hypothesis of the paper is that the use of soft-clustering with our prediction approach is adequate to deal with recommendation and clustering in not traditional CF RS datasets, such as those containing scientific documentation. The secondary hypothesis leads us to predict that there will be optimal overlapping values that provide a balance between quality of clustering and quality of recommendation. Our hypothesis is motivated in two aspects:

1) In the scientific documentation datasets, each item can be classified naturally into several categories, which justifies the use of soft-clustering; e.g.: ‘‘machine learning’’ is a topic that may appear, to a greater or lesser extent, in very different fields, such as: natural language processing, artificial vision, speech processing, recommender systems, etc.
2) The scientific documentation datasets contain less regular and more complex information than their equivalent traditional datasets: mainly their non-regular frequency distributions and their wide range of cardinalities (ratings), which motivates the use of soft-clustering techniques because they provide greater flexibility and parametric adaptability than hard-clustering techniques.

Based on the explained motivations and the established hypothesis, the main objectives of this paper are:

1) To provide a new prediction approach for probabilistic soft clustering methods.
2) To perform experiments [6] that allow to compare: a) Results obtained using SD4AI and another traditional CF dataset of similar size, and b) Results obtained using our approach in two soft clustering methods: the memory-based FCM [19], [20] and the model-based BNMF [10].
3) To find overlapping values that provide an optimal balance between quality of clustering and quality of recommendation.
4) To determine the quality with which the recommendation and the clustering of not traditional datasets, such as SD4AI, can be addressed.

Once the motivation, hypothesis and objectives of the paper have been established, we turn to show the state of the art in the research fields directly related to our proposal. The use of RS clustering [19], [21], [22] information can be innovative and very useful; traditionally, this information has been used to pursue the usual goal: improving RS accuracy [1]. However, recently, two important RS aims have arisen in which this information is adequate: 1) Recommendations explanation [23], and 2) RS visualization [24]. There is an innovative use, of increasing importance, whose approach can be considered paradigmatic: data analytics. We can consider clustering analysis as the next step of two previous traditional fields in CF RS: 1) Items recommendation to each user [25], and 2) Items recommendation to groups of users (e.g. friends who go, together, to the cinema) [26], [27]. Using data analytics approach, instead of making recommendations to users (about existing products), recommendations are made to the companies (about not existing products). In this way, companies will know which product types are more suitable to the different client profiles of their RS. This new CF tool will serve both users and companies: 1) Increasing the offered products adequacy, 2) Providing strategic information to companies, 3) Improving economic benefits.

In recent years, more and more publications have emerged covering beyond accuracy objectives [28]. We want to accomplish an important beyond accuracy goal: obtaining RS clusters with the best possible quality, taking into account the restrictions of the CF datasets. CF datasets nature is highly sparse, particularly those based on explicit ratings; users can only vote for a small proportion of the large number of available items. The highly sparse nature of the datasets [29], [30] causes three important consequences in the clustering process: 1) Traditional clustering methods are not accurate,

2) The best metrics to measure elements distance are not the usual ones, and 3) MF methods are suitable machine learning candidates for clustering, due to their accuracy in the RS prediction and recommendation.

It is possible to formulate clustering as a matrix decomposition problem [11], [29]. According to [31] and [32] MF has important advantages when used as a clustering method: 1) It can model widely varying data distributions due to the flexibility of matrix factorization [33], [34]; 2) It is able to perform simultaneous clustering of the rows (users) and the columns (items) of the input data matrix; and 3) It can simultaneously achieve both hard and soft clustering.

The following are some representative papers in the field of CF clustering: Birtolo and Ronca [20] propose two clustering-based CF algorithms: Item-based fuzzy clustering and trust-aware clustering; they obtain an increased value of coverage without affecting recommendation quality. Zhang *et al.* [35] optimize the standard MF by integrating the user clustering regularization term; they report improvements in the recommendation accuracy, compared with standard algorithms. To dimension the number of clusters (K) is a process that requires experience, knowledge of the data and a trial and error mechanism to choose the most appropriate values. The correct choice of this parameter determines the quality of the resulting clusters, as well as the predictions and recommendations made; [36] dynamically sets the parameter: with more data coming in, the incremental clustering algorithm determines whether to increase the number of clusters or merging the existing clusters. Authors report encouraging prediction accuracy. Wu *et al.* [37] use a co-clustering method to divide the raw rating matrix into clusters, and then it employs NMF to make improved predictions of unknown ratings.

NMF [38] can model widely varying data distributions due to the flexibility of MF as compared to the rigid spherical clusters that the K-means clustering objective function attempts to capture. When the data distribution is far from a spherical clustering, NMF may have advantages. NMF scalability can be effectively addressed through different schemes: Shrinking, partitioning [39], incremental [40] and parallel [41], making it possible to perform clustering of big data CF matrices.

The RS datasets nature leads us towards a soft clustering approach [42]–[44]: On the one hand, RS users can be interested in various types of items (e.g. gadgets, sports and fashion products). Items can be simultaneously framed in several types of features (e.g. science-fiction and horror film); in this way, it is not appropriate to classify RS users and items in a hard clustering rigid manner. It is preferable to provide the soft clustering approaches flexibility degree, where each user and each item are probabilistically defined according to several hidden characteristics (e.g. movie: 0.0 humor, 0.7 action, 0.7 horror, 0.9 science fiction, 0.0 musical ...).

As the proposed method, we have designed a new prediction approach for probabilistic soft clustering algorithms. As soft clustering methods, we have selected BNMF [9] and FCM [12], [14], [15], [17], [19], [20]. BNMF [10] is a recently published Bayesian Non-Negative Matrix Factorization [45]–[49]. This method provides a parameter, alpha, which allows to probabilistically vary the dispersion of the hidden factors of each user or item, and therefore the membership degree of each user or item to each cluster. It is based on factorizing the ratings matrix into two non-negative matrices whose components have an understandable probabilistic meaning.

Current soft clustering relevant papers are: [42], where authors propose a new fuzzy hierarchical clustering algorithm, applied to CF; they claim accuracy improvements and user-product joint groups detection. Siminski [50] presents a fuzzy clustering algorithm designed to filter low importance datasets attributes. Vidhya and Geetha [43] review clustering based on rough set theory. Deng *et al.* [44] provide a survey on soft subspace clustering, where clusters are based in association with subspaces of high-dimensional spaces. Hai-Peng *et al.* [19] address the problem of automatically find the optimal number of clusters. Koohi and Kiani [17] show the superiority of fuzzy clustering approaches, compared to the traditional ones when applied to RS.

To measure the validity of each soft clustering is important [51]. We can compare the proposed method results and the baseline ones by using several clustering quality measures. Within-cluster (compactness) and between-cluster (separation) are popular and intuitive quality measures; Muranishi *et al.* [52] present a fuzzy validity criterion which identifies compact and separate fuzzy partitions. Xie and Beni [53] and Grekousis and Thomas [54] cluster validity index evaluates FCM clustering quality considering geometrical features of clusters; cluster compactness and separateness are measured by using intra-cluster deviations and inter-cluster distance. The high sparsity [30] levels presented by the CF RS datasets make it much more complex to obtain accurate clustering; we can not ignore that we are working in a field where predictions and recommendations are a fundamental objective. In this way, it is appropriate to incorporate some prediction accuracy [1], [55] measure to the clustering quality measures set.

The rest of the paper is structured as follows: Section II formalizes the proposed method and summarizes the algorithms soft clustering BNMF y FCM, section III introduces the experiments designed for the paper, section IV explains the obtained results. Finally, section V summarizes the conclusions of the paper and it proposes promising future works.

II. PROPOSED METHOD

The proposed method named ***Soft Predict (SP)*** combines the information of the user's membership matrix with the information of the centroids of each cluster. The main idea of the method is that it executes the recommendation to an active user by analyzing the information with respect to all the groups and not only to the group to which it most likely belongs. The equation 6 formalizes this approach.

- The membership matrix contains the probability of belonging to each group to each cluster.

TABLE 1. Parameters and measures.

Name	Descriptions
x	User (or paper)
n	$\#x$ in dataset
y	Neighbor user (or paper) of x
i	Item (or topic)
α	Overlapping parameter BNMF, $\alpha \in (0, 1]$ $\alpha \approx 0$, the lower degree of overlapping. $\alpha = 1$, the upper degree of overlapping.
β	Parameter of control of evidence so that an item like a group (or topic exists in a group) in BNMF, $\beta > 1$ $\beta \gg 1$, the algorithm requires more evidence to assign x to a cluster
$\lambda_{x,i,j}$	Variable BNMF that represents the probability that users x in cluster j like the item i (or probability that papers x in cluster j contains the topic i)
$\gamma_{x,j}$	Auxiliary variable BNMF
$\epsilon^+_{i,j}, \epsilon^-_{i,j}$	Auxiliary variables BNMF
m	Overlapping parameter FCM, $m > 1$ $m \approx 1$, the lower degree of overlapping. $m \gg 1$, the upper degree of overlapping.
$r_{x,i}$	Rating (or cardinality) of x to i
$r_{y,i}$	Rating (or cardinality) of y to i
$M\left(r_{x,i}\right)$	Ratings (or cardinalities) matrix
$M\left(r^*_{x,i}\right)$	Ratings (or cardinalities) normalized matrix
K	#clusters (#factors in BNMF)
K^*	Optimum number of clusters
$a_{x,j}$	Membership matrix. Probability that x belongs to cluster j
j	Cluster index (factor index in BNMF)
C_j	Centroid of cluster j
F_j	$\#x$ in cluster j
ℓ_j	#neighbors in cluster j
$P_{x,i}$	Predicction of x to i
$\bar{r}_x$	Rating (or cardinality) average of x
$\bar{r}_y$	Rating (or cardinality) average of y

- A centroid contains the preferences (or cardinalities) of a cluster on each item (or topic).

We present the table of parameters and measures (Table 1) used in the formalizations of our proposed method made in the paper.

Algorithms 1 and 2 summarize our original version of the soft clustering methods BNMF and FCM. For details about BNMF and FCM clustering phase see [10], [14] respectively. We have incorporated our *Soft Predict (SP)* approach to probabilistic soft clustering methods. In BNMF our approach corresponds to step 16, while in FCM it corresponds to step 9. Originally BNMF [10] does not generate the output of centroids, it just provides the membership matrix. We have incorporated the calculation of centroids into BNMF in order to carry out our prediction approach. The calculation of a centroid is done by the weighted aggregation of the r_x preferences of the users of cluster j. The weights are the belonging probabilities of the users to said cluster $a_{x,j}$.

Additionally, we propose a soft clustering strategy (algorithm 3) to collaborative filtering recommender systems. This strategy determines the best soft parameter values based on the quality of clustering (minimization of XB). In an iterative way, MAE and XB are measured as the number of clusters increases. The algorithm finishes when any cluster does not contain users (or papers); this is an indicator that no more clusters are required. The algorithm generates, as output, the *Overlapping* parameter and the number of clusters K with which the lowest XB has been obtained.

Algorithm 1 BNMF

Input: M, K, α, β, *iterations*
Output: $a_{x,j}$, C_j and $P_{x,i}$

1: Initialize randomly $\gamma_{x,j}$, $\epsilon^+_{i,j}$ and $\epsilon^-_{i,j}$
2: $iter = 0$
3: **repeat**
4:   **foreach** x:
5:     **foreach** i rated by (or exists in) x:
6:       **foreach** factor j: update $\lambda_{x,i,j}$ (eq. 9)
7:   **foreach** x:
8:     **foreach** factor j: update $\gamma_{x,j}$ (eq. 1)
9:   **foreach** i rated by (or exists in) x:
10:     **foreach** factor j: $\epsilon^+_{i,j}$ (eq. 2)
11:     **foreach** factor j: $\epsilon^-_{i,j}$ (eq. 3)
12:   $iter = iter + 1$
13: **until** changes are not significant OR $iter = iterations$
14: compute membership matrix $a_{x,j}$ (eq. 4)
15: **for each** cluster j: compute the updated cluster center C_j (eq. 5)
16: compute *Soft Predict* $P_{x,i}$ (eq. 6)
17: **return** $a_{x,j}$, C_j and $P_{x,i}$

Algorithm 2 FCM

Input: M, K, m, *iterations*
Output: $a_{x,j}$, C_j and $P_{x,i}$

1: C=**Centroid initialization randomly**
2: $iter = 0$
3: **repeat**
4:   **foreach** user (or paper) x
5:     compute membership function $a_{x,j}$ (eq. 7)
6:     compute the updated cluster center C_j (eq. 8)
7:   $iter = iter + 1$
8: **until** changes are not significant OR $iter = iterations$
9: compute *Soft Predict* $P_{x,i}$ (eq. 6)
10: **return** $a_{x,j}$, C_j and $P_{x,i}$

III. EXPERIMENTS DESIGN

This section explains the experiments design: chosen datasets, soft clustering tested methods, quality measures, parameter values, etc. Experiments are performed using cross-validation. We compared recently published approaches to FCM [12], [14], [15], [17], [19], [20] with our prediction approach *FCM-SP*. Our approach applied to BNMF [10] yields similar quality results, since this model does not admit significant improvements.

The chosen datasets are the public *Movielens 1M* [3] and the *SD4AI* [18] ones. The soft clustering methods are *BNMF* [10] and our proposed method *FCM*. Finally, we test diverse clustering quality measures (*F-Partition*, *cohesion*,

TABLE 2. Datasets size and composition.

SD4AI		*Movielens 1M*	
# papers	14,143	# users	6,240
# topics	18,502	# movies	3,952
# cardinalities	1,389,094	# ratings	1,000,209
sparsity	99.47%	sparsity	95.94%
range	[1..160] step 0.25	range	{1, 2, 3, 4, 5}

separation and *Xie and Beni index*), and the *MAE* prediction quality measure.

To run the designed experiments, we have chosen two open datasets: *Movielens 1M* and *SD4AI*. These two RS collaborative filtering datasets are very different, and this circumstance will help us to compare the performance of the clustering methods: *BNMF* and *FCM*. *Movielens* is a classical collaborative filtering dataset, containing votes casted by users to movies. *SD4AI* is a scientific documentation datamined dataset containing cardinalities of topics from each paper. Table 2 shows the main parameter values of both datasets: they have a similar number of ratings, but *SD4AI* holds much more papers and topics than *Movielens* users and movies does; consequently, *SD4AI* is sparser than *Movielens*. On the other hand, the ranges of votes/cardinalities of the tested datasets are radically different: to obtain accurate *SD4AI* predictions will be more difficult than getting it using *Movielens*.

$$\gamma_{x,j} = \alpha + \sum_{\{i|r_{x,i} \neq \bullet\}} \lambda_{x,i,j} \tag{1}$$

$$\epsilon^{+}_{i,j} = \beta + \sum_{\{x|r_{x,i} \neq \bullet\}} \lambda_{x,i,j} \cdot r^{+}_{x,i} \tag{2}$$

$$\epsilon^{-}_{i,j} = \beta + \sum_{\{x|r_{x,i} \neq \bullet\}} \lambda_{x,i,j} \cdot r^{-}_{x,i} \tag{3}$$

$$a_{x,j} = \frac{\gamma_{x,j}}{\gamma_{x,1} + \ldots + \gamma_{x,K}} \tag{4}$$

$$C_j = \frac{\sum_{x=1}^{n} \left(a_{x,j}\right) r_x}{\sum_{x=1}^{n} \left(a_{x,j}\right)} | r_{x,i} \neq \bullet \} \tag{5}$$

where: C_j is the centroid of cluster j and r_x are the ratings (or cardinalities) of x

$$P_{x,i} = \sum_{j=1 \ldots K} a_{x,j} \cdot C_{j,i} \tag{6}$$

$$a_{x,j} = \frac{1}{\sum_{l=1}^{K} \left(\frac{||r_x - C_j||^2}{||r_x - C_l||^2} \right)^{2(m-1)}} \tag{7}$$

$$C_j = \frac{\sum_{x=1}^{n} \left(a_{x,j}\right)^m r_x}{\sum_{x=1}^{n} \left(a_{x,j}\right)^m} \tag{8}$$

Figure 2 shows the frequency distribution of the *Movielens 1M* dataset ratings (left graph) and the frequency distribution of the *SD4AI* dataset cardinalities (right graph). We can observe that both datasets provide very different distributions:

Algorithm 3 Soft Clustering Strategy

Input: $M(r_{x,i})$,
Output: K^*, *Overlapping*

select *RS_soft-clustering algorithm* (algorithm 1 or algorithm 2)
$K = 2$, $XB^* \approx \text{inf}$, $K^* = 1$
repeat
  determinate *Overlapping* parameter (m in FCM, α in BNMF)
  run *RS_soft_clustering algorithm*
  compute *XB*
  compute *MAE*
  if $XB < XB^*$ **then**
    $XB^* = XB$
    $K^* = K$
  end if
  $K = K + 1$
until $F_j = 0$ in any cluster j
return K^*, *Overlappingparameter*

TABLE 3. Cross-validation values used in the experiments.

General parameters	
Testing-Items%	20%
Testing-Users%	20%
Training-Items%	80%
Training-Users%	80%
#clusters (K)	Movielens: {2, 5, 10, 15, 20} SD4AI: {2, 5, 10, 20, 30, 40, 50, 60}
FCM m soft parameter	Movielens: {1.01, 1.03, 1.07, 1.1, 1.25} SD4AI: {1.1, 1.25, 1.5, 1.75, 2}
BNMF α soft parameter	[0,1..1] step 0.1

whereas *Movielens* shows a normal-like distribution, *SD4AI* does not. The broad range of cardinalities in *SD4AI* and their distribution makes prediction a challenging task when we use this dataset. Experiments have been run using cross-validation, and each experiment set of parameter values has been tailored to the size and the nature of each dataset. The soft methods overlapping values (m in *FCM* and α in *BNMF*) have been chosen to comprise the whole hard to soft operating modes of *FCM* and *BNMF*. Table 3 contains the set of cross validation and parameters values.

The quality measures we use in the experiments are:

- *F-Partition coefficient*: this parameter measures the amount of overlap between clusters. If F is 1 there is no membership sharing (extreme hard clustering). If F is 0 there is a total membership sharing between clusters (extreme soft clustering). Most real situations require a balance in the membership sharing between clusters.
- *Soft compactness (soft cohesion)*: it measures the distances from each item j (X_j) to each cluster i (V_i). Since we are using soft clustering methods, we weight each distance by using $\mu_{i,j}$: the probability of item j to belong to the cluster i. The soft compactness can be formalized

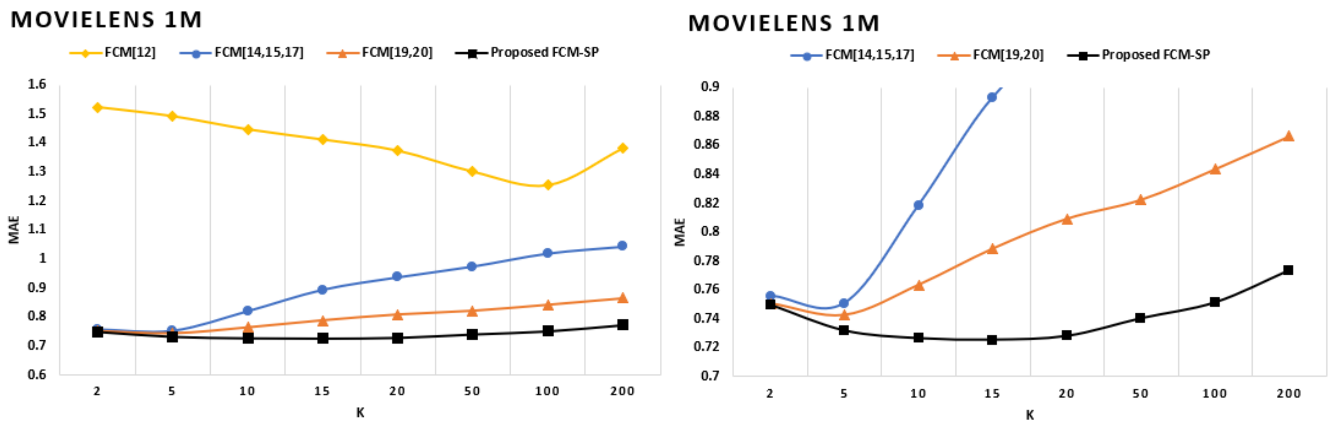

FIGURE 1. ***Mean Absolute Error (MAE)*** **results using several state of the art baselines** ***FCM*** **methods and the proposed** ***FCM-SP*** **method. The** ***Movielens*** **dataset is applied.** ***y*****-axis:** ***MAE*** **results;** ***x*****-axis: number of clusters (*****K*****). Lower values are the better ones.**

as:

$$compactness = \frac{\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{i,j}^{2} \parallel V_i - X_j \parallel^2}{n}$$

where n is the number of items, and c is the number of clusters. Low soft compactness values are better, since it means that items belonging to each cluster are near.

- *Soft separation*: it measures the minimal distance between any pair of the cluster centroids. The soft separation can be formalized as:

$$d_{min} = min_{i,j} \parallel V_i - V_j \parallel$$

where i and j are any pair of cluster centroids. High soft separation values are better, since it means that clusters are more separated.

- *Xie and Beni index (XB)*: it just makes the *compactness* divided by *separation*. *XB* provides a unified value for the above clustering quality measures. Its drawback is that it loses the details of both *compactness* and *separation* values.
- *Mean Absolute Error (MAE)*: this is not a clustering quality measure; it is a collaborative filtering prediction quality measure. The *MAE* can be formalized as:

$$MAE = \frac{1}{|\{r_{u,i}|r_{u,i} \neq \bullet\}|}\sum_{u\in U}\sum_{i\in I}|p_{u,i} - r_{u,i}|$$

where $r_{u,i}$ is the user u rating to the item i. $p_{u,i}$ is the prediction of $r_{u,i}$, and $\{r_{u,i}|r_{u,i} \neq \bullet\}$ is the set of casted ratings. Low MAE values are better, since it means that prediction errors are lower.

IV. RESULTS

The results' subsection is split in 6 figures and their explanations. Each figure shows a quality result. Figure 1 (left side) shows the best quality of predictions obtained when different FCM clustering experiments [12], [14], [15], [17], [19], [20] are run (best results are obtained by using m=1.1). FCM [12] presents the worst *MAE* results because its prediction approach based on the K more similar centroids tends to generate general prediction values for each group. The results of *MAE* in the FCM [14], [15], [17]baselines get worse as the value of K increases, similarly happens in FCM baselines [19], [20]. Note that our proposed *FCM-SP* method has better results in terms of *MAE* with respect to the state of the art FCM methods. The Figure 1 (right side) shows the results without FCM baseline [12] to appreciate the difference among the other baselines FCM and our proposed method. This improvement is more significant with a high K. Therefore, our *FCM-SP* method has been selected for the rest of experiments. From here on, our method will be named only as FCM.

$$\begin{aligned}\lambda'_{x,i,j} &= exp\left(\Psi(\gamma_{x,j}) + r^{+}_{x,i}\cdot\Psi(\epsilon^{+}_{i,j}) + r^{-}_{x,i}\cdot\Psi(\epsilon^{-}_{i,j}) - R\cdot\Psi(\epsilon^{+}_{i,j}+\epsilon^{-}_{i,j})\right)\\ \lambda_{x,i,j} &= \frac{\lambda'_{x,i,j}}{\lambda'_{x,i,1}+\ldots+\lambda'_{x,i,K}}\end{aligned} \tag{9}$$

where:

- Ψ is the digamma function defined as the logarithmic derivative of the gamma function: $\Psi(x) = (ln\,\Gamma(x))' = \frac{\Gamma'(x)}{\Gamma(x)}$
- R = max ($r_{x,i}$) - interval_rating_range (or interval_cardinality_range)
- $r^{+}_{x,i} = \rho_{x,i} = R\cdot r^{*}_{x,i}$
- $r^{-}_{x,i} = R - \rho_{x,i} = R\cdot(1 - r^{*}_{x,i})$

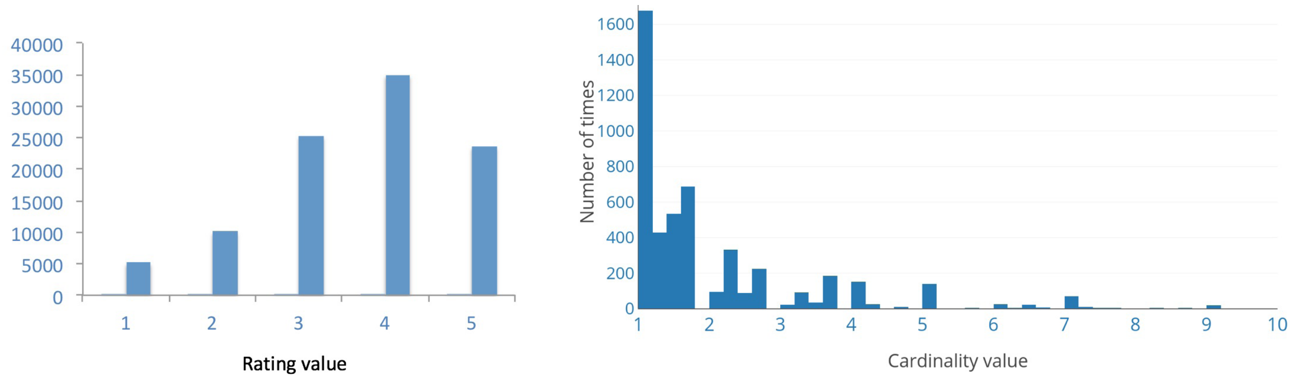

FIGURE 2. Frequency distribution of the *SD4AI* dataset cardinalities (right) and the *Movielens 1M* dataset ratings (left). *X*-axis: most representative cardinality values and rating values. *Y*-axis: frequency of each cardinality (number of times this cardinality appears in the dataset), and frequency of each rating (number of times this rating appears in the dataset).

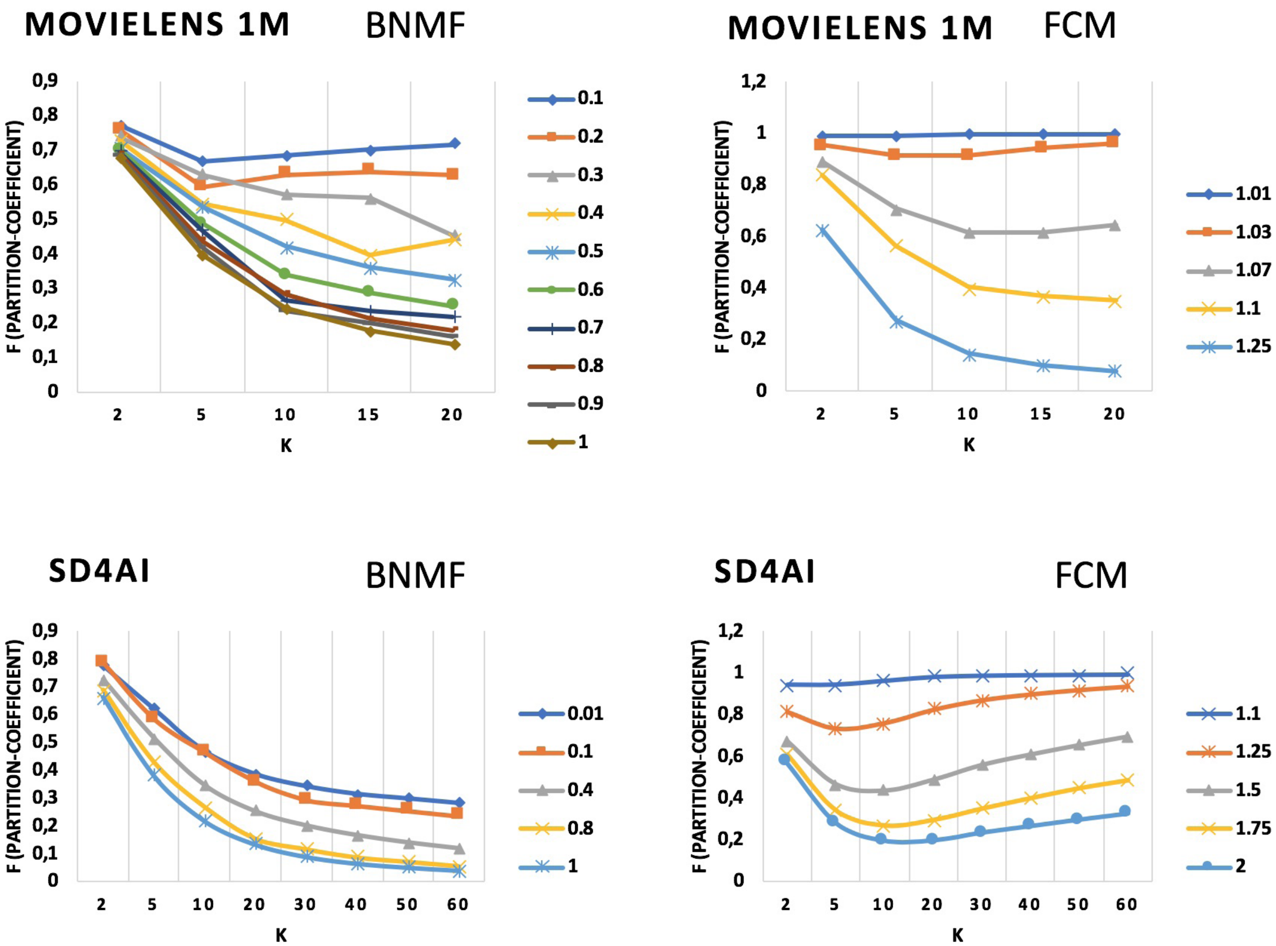

FIGURE 3. *F-partition* coefficient results using *BNMF* (left graphs) and *FCM* (right graphs) methods. The *Movielens* (top graphs) and *SD4AI* (bottom graphs) datasets are applied. *y*-axis: *F* partition results; *x*-axis: number of clusters (*K*). Cross validation values stated in Table 3. Low *F* values means soft clustering behavior, whereas high *F* values means hard clustering behavior.

The next figures show results obtained from the baseline *Movielens* and the tested *SD4AI* datasets. Both *BNMF* and *FCM* methods are applied to the above datasets. The quality results are: *F-Partition*, *compactness*, *separation*, *MAE* and *XB versus MAE*.

Figure 3 shows the *F-partition* coefficient. It let us to make a fine tuning of the soft clustering parameters for both *BNMF* and *FCM* methods. As an example, when we use the *FCM* clustering method on the *Movielens* dataset (top-right graph in Figure 3), the $m = 1.01$ value provides us an extreme

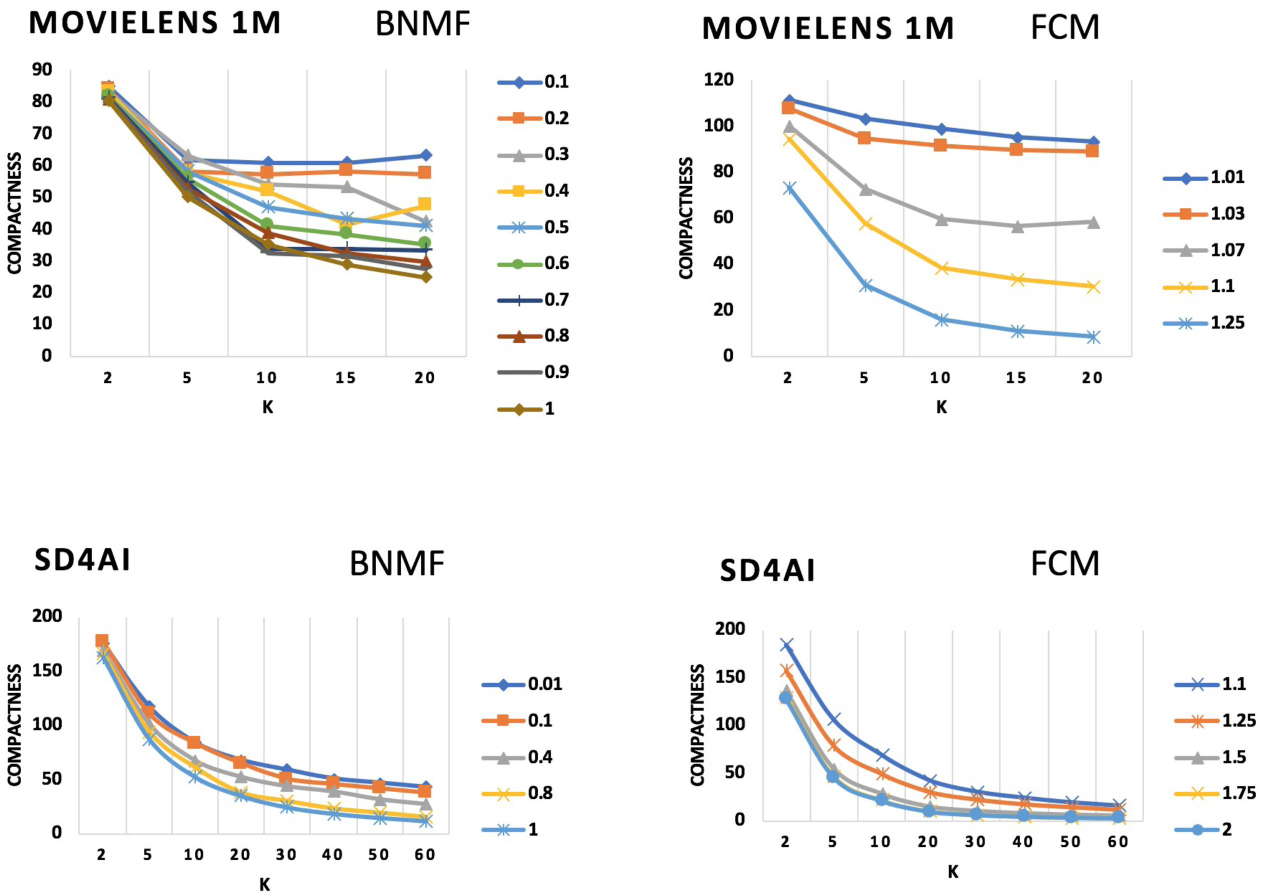

FIGURE 4. ***Compactness (cohesion)*** **results using** ***BNMF*** **(left graphs) and** ***FCM*** **(right graphs) methods. The** ***Movielens*** **(top graphs) and** ***SD4AI*** **(bottom graphs) datasets are applied.** ***y*****-axis: soft compactness results;** ***x*****-axis: number of clusters (*K*). Cross validation values stated in Table 3. Lower values are the better ones.**

hard clustering behavior, and the $m = 1.25$ provides us an extreme soft clustering behavior on K values bigger than 15. This means that experiments can be delimited to this range of m values. When the *SD4AI* is used (bottom-right graph in Figure 3), the *FCM* m parameter needs to get a 2.0 value to reach a high soft clustering behavior; this is due to the particularly high sparsity of this dataset and to its wide range of cardinality values (rating values). From Figure 3 (left graphs) we can observe a uniform behavior relating the *BNMF* parameter values and the obtained *F-Partition* coefficient results; this means that we should use the complete α (0..1] range to test *BNMF* results, both in *Movielens* and *SD4AI*.

Figure 4 shows the *compactness* clustering quality measure results. As we explained in section III, the soft version is tested. Clustering is better when the items belonging to each cluster are near, so the lower the *compactness* value the better the clustering quality. Results show absolute values; for this reason, it is not adequate to compare *SD4AI* and *Movielens* ranges, but it is very interesting to analyze the different graph evolutions. As expected, the greater the value of K the better the clustering, and therefore the better (lower) *compactness* results are achieved.

From Figure 4 we can state that the *FCM* method provides better clustering results than the *BNMF* one, but this is true when we select their soft behavior (the highest α and m values). This observation is consistent with the hypothesis of the paper. Additionally, we can see that the *Movielens* dataset is more sensitive to the soft/hard clustering variations of the algorithms: *compactness* results are more homogeneous in the *SD4AI* dataset.

Figure 5 shows the complementary quality measure to the Figure 4 one (*compactness*): Figure 5 shows the existing soft *separation* from the set of K clusters. It is adequate that clusters are separated as much as possible to minimize the number of cases in which it is risky to assign an item to one cluster or another. In this way, high *separation* values are better. From graphs in Figure 5 we can state the following regards:

- As expected, *compactness* and *separation* quality measures are opposed to each other: the bigger the value of K, the better the *compactness* results and the worse the *separation* values: if the number of

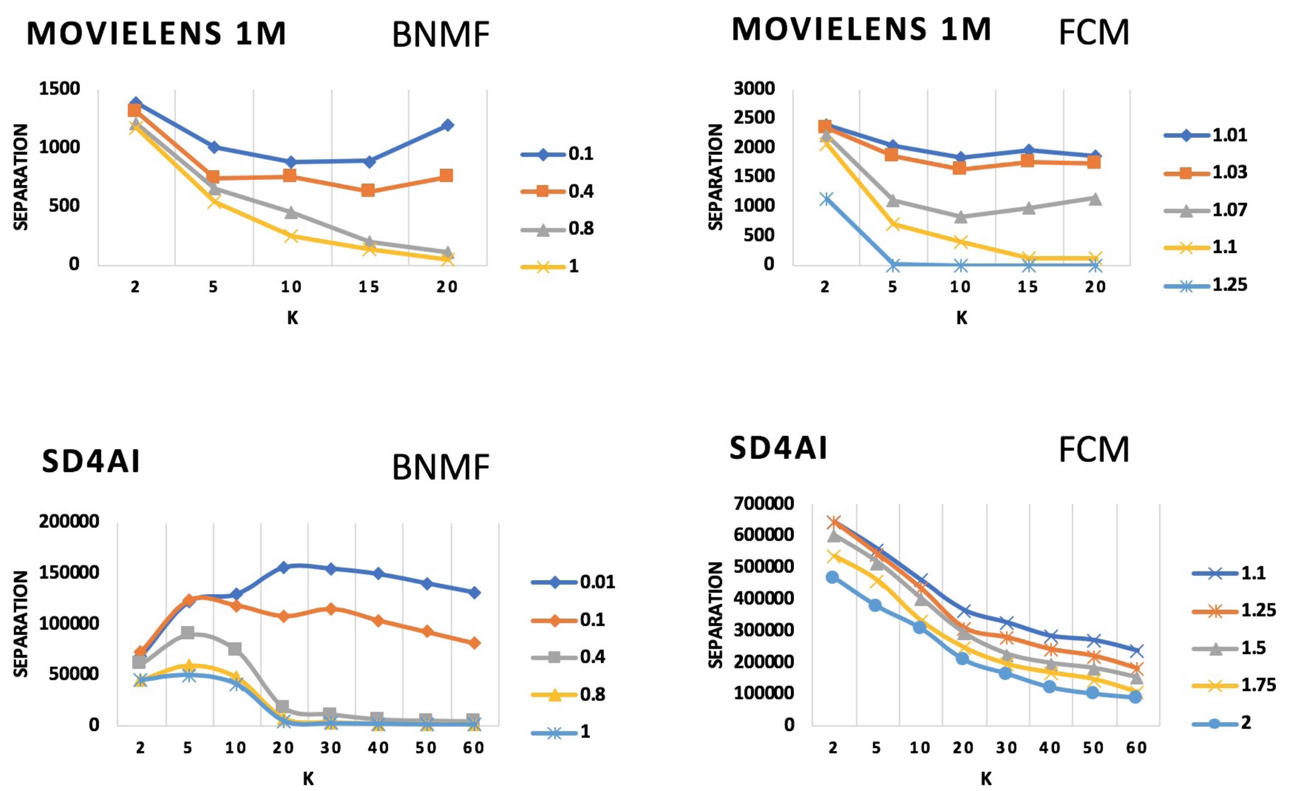

FIGURE 5. ***Separation*** **results using** ***BNMF*** **(left graphs) and** ***FCM*** **(right graphs) methods. The** ***Movielens*** **(top graphs) and** ***SD4AI*** **(bottom graphs) datasets are applied.** ***y*****-axis: soft separation results;** ***x*****-axis: number of clusters (*K*). Cross validation values stated in Table 3. Higher values are the better ones.**

- clusters increases the separation between them tends to decrease.
- While the soft values of the *FCM* and *BNMF* methods improve the clustering *compactness*, those same soft values worsen the *separation* results. This is a very important aspect, because it leads us to the choice of a balanced operating mode between soft and hard clustering. In Figure 7 we will unify *compactness* and *separation* quality measures, using *XB* (see section III).

Since we are testing recommender systems collaborative filtering datasets, it is rational to use some collaborative filtering quality measure. We have chosen the classical mean prediction error one: *MAE*. Figure 6 shows the quality of predictions obtained when different clustering experiments are run. Note that the absolute *MAE* values vary between *SD4AI* and *Movielens*, because these datasets have very different rating ranges. The main observations that are derived from the graphs contained in Figure 6 are:

- The balanced hard/soft clustering parameter values are those that provide better predictions (lower errors). This choice of soft clustering equilibrium is compatible with that obtained with *compactness* & *separation* measures.
- As expected, the optimal K is lower in *Movielens* than in *SD4AI*, due to the very nature of each dataset: there is less variety in the preferences of the films than in the research topics and their relationships.
- *BNMF* provides better prediction values when applied to the *Movielens* dataset; conversely, *FCM* provides better prediction values when applied to the *SD4AI* dataset. An explanation for this behavior is the design of *BNMF*, which assumes a limited range of observation values; *Movielens* fulfills this premise, while *SD4AI* does not. This is an important aspect to keep in mind when the clustering process is intended to be used in obtaining recommendations.

Figure 7 is particularly interesting, since it shows both a mix of the preceding figures and a balance between clustering quality and prediction quality. Figure 7 shows graphs where clustering quality is condensed (*XB* quality measure) and where prediction quality is displayed. Both clustering quality and prediction quality values are normalized in order to be compared in the same range [0..1]. *Movielens* experiments are made using the fixed number of clusters $K = 15$, whereas *SD4AI* experiments are run using a $K = 30$ value; both values are taking from Figure 6, selecting the K values where results are better (lower). The graphs' x axis shows the α *BNMF* and the m *FCM overlapping* values (soft/hard balance parameters). Both the *MAE* and the *XB* quality measures provide their better results when values are low. From Figure 7 we can state:

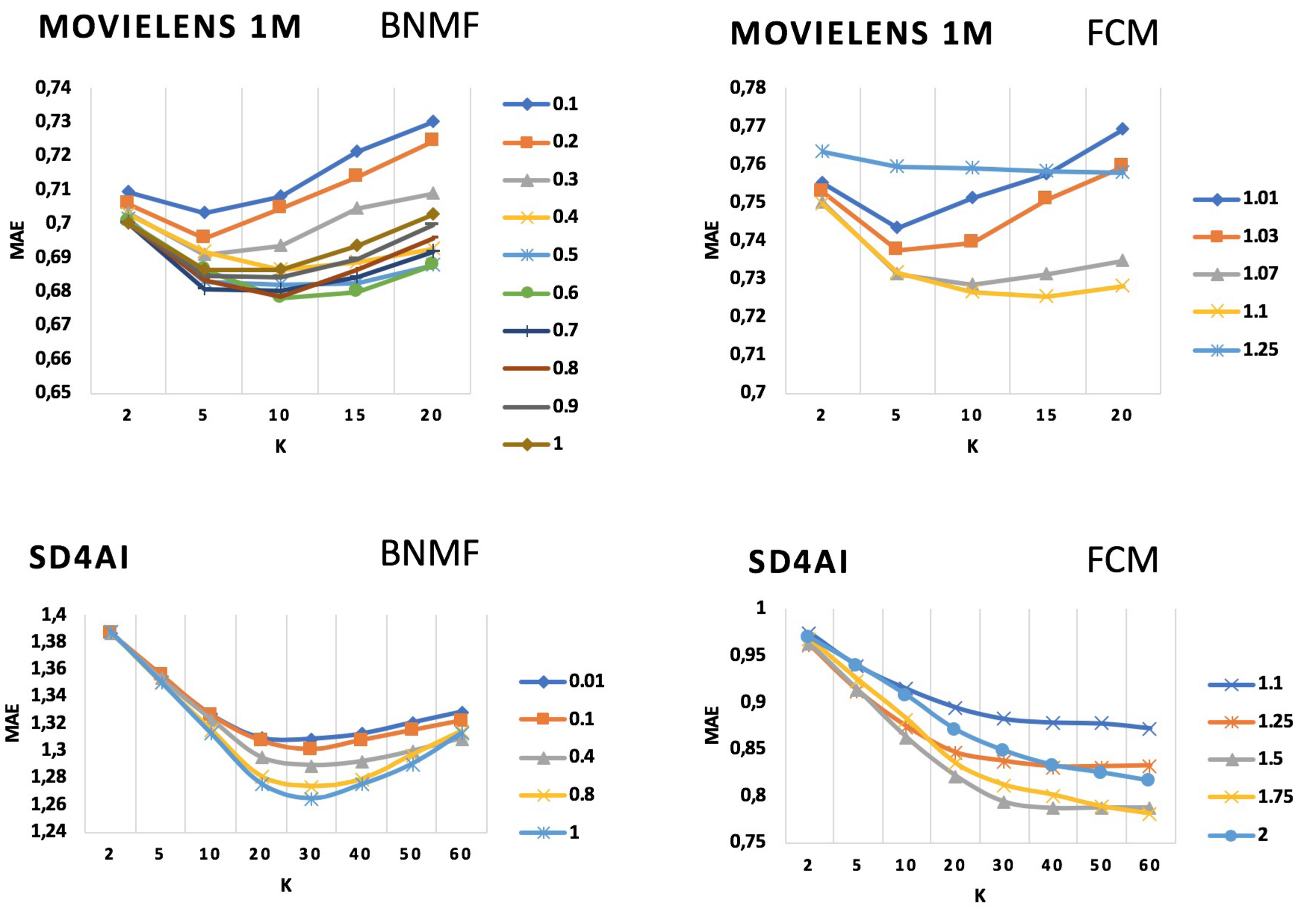

FIGURE 6. ***Mean Absolute Error (MAE)* results using *BNMF* (left graphs) and *FCM* (right graphs) methods. The *Movielens* (top graphs) and *SD4AI* (bottom graphs) datasets are applied. *y*-axis: *MAE* results; *x*-axis: number of clusters (*K*). Cross validation values stated in Table 3. Lower values are the better ones.**

- There are intersection points between the curves; they represent adequate balances for clustering quality and prediction quality. These intersection points provide us the optimum *overlapping* values for both *FCM* and *BNMF* methods; e.g.: $\alpha = 0.36$ in the *BNMF* method applied to the *SD4AI* dataset, or the $m = 1.5$ in the *FCM* method applied to the *SD4AI* dataset.
- Optimum *overlapping* values are balanced: they provide a balanced soft/hard clustering behavior; e.g.: *BNMF* $\alpha = 0.36$ & $\alpha = 0.55$ in the [0..1] range. *FCM* $m = 1.5$ in the [1,1..2] range & $m = 1.1$ in the [1..1,25] range.
- *FCM* provides better performance than *BNMF* for both datasets (their y-axis values are lower).
- *Movielens* provides better performance than *SD4AI* for both soft clustering methods (their y-axis values are lower). This is the expected behavior due to the *SD4AI* higher sparsity and wider cardinality (rating) range.

V. CONCLUSIONS

The most relevant conclusion obtained from the experiment results is the confirmation of the paper's hypothesis: our proposed approach improves the prediction quality results compared to other recently published FCM soft clustering methods, soft-clustering methods are suitable for processing not traditional CF RS datasets, such as *SD4AI*. In addition, it is verified that quality of recommendation and quality of clustering are opposing objectives, and that we can find *overlapping* values that optimize both objectives. Intermediate *overlapping* values (between soft and hard clustering) offer balanced recommendation and clustering quality results. This confirms our secondary hypothesis: the convenience of using probabilistic clustering methods. By using soft-clustering methods we can fine-tune their parameters to find optimal values.

The scientific documentation datasets represent a challenge in the CF RS field, due to their especially complex characteristics: irregular frequency distributions, broad range of rating values and high sparsity. The use of soft-clustering methods has proven to be an effective tool to obtain adequate results under these extreme conditions.

Results of this paper open some promising future works: a) Creation of new not traditional CF datasets, based on data mining, b) Introduction of the recommender systems area in the scientific documentation field, c) Design of new

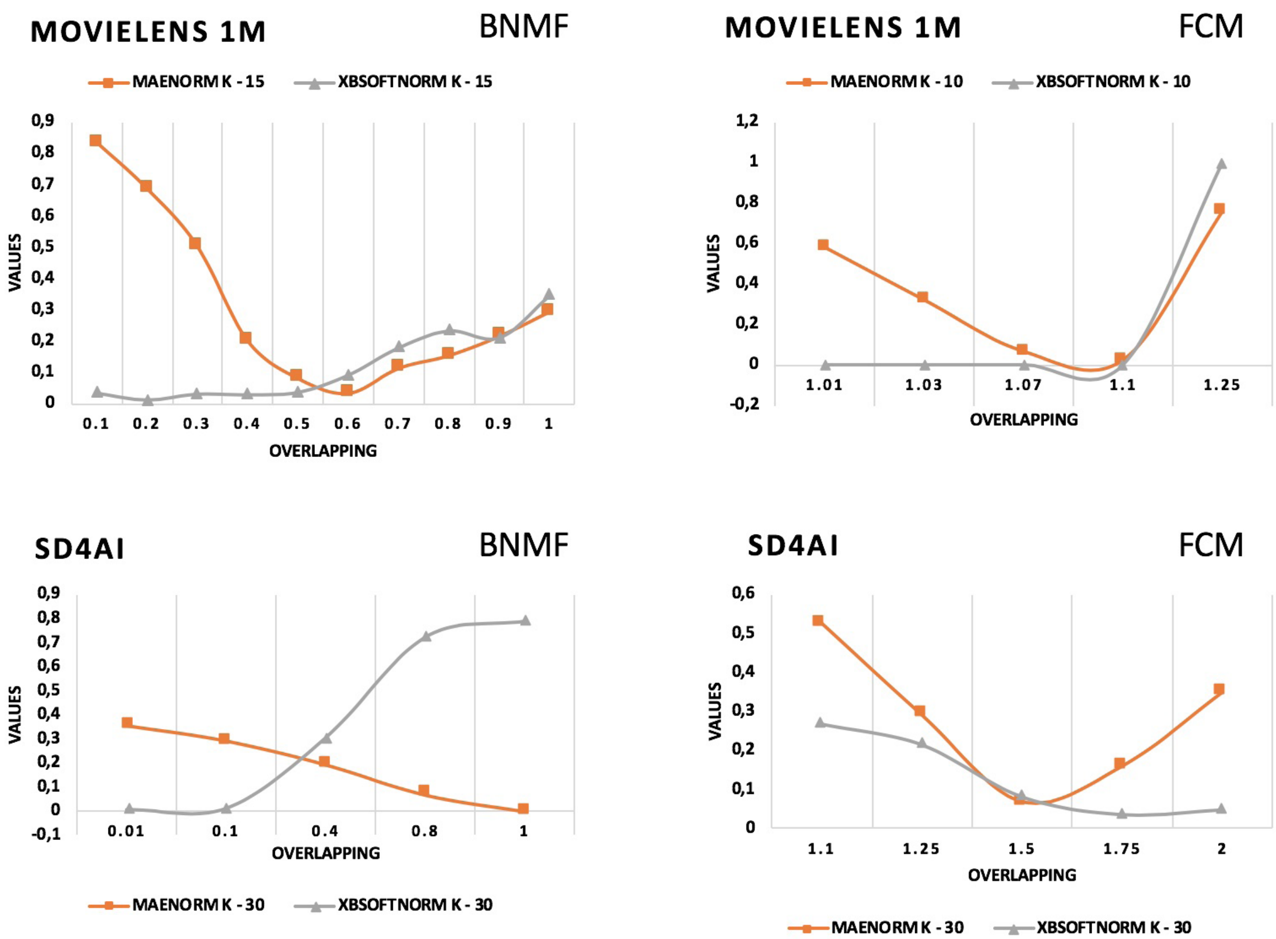

FIGURE 7. ***Mean Absolute Error (MAE) versus XB*** **results using** ***BNMF*** **(left graphs) and** ***FCM*** **(right graphs) methods. The** ***Movielens*** **(top graphs) and** ***SD4AI*** **(bottom graphs) datasets are applied.** ***y*-axis: *normalized MAE & normalized XB*** **results;** ***x*-axis: *overlapping*** **values (soft/hard balance). Cross validation values stated in Table 3. Lower values are the better ones.**

model-based soft-clustering algorithms and d) Future recommender systems applications where the size and distribution of the ratings are more challenging than in the current datasets; these applications will be fed, in many cases, by IoT data.

REFERENCES

[1] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowl.-Based Syst.*, vol. 46, pp. 109–132, Jul. 2013.

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

[3] GroupLens. (2000). *Movielens Dataset*. [Online]. Available: https://grouplens.org/datasets/movielens/

[4] T. Bertin-Mahieux, D. P. Ellis, B. Whitman, and P. Lamere, "The million song dataset," in *Proc. 12th Int. Conf. Music Inf. Retr. (ISMIR)*, 2011, p. 10.

[5] Book-Crossing. (2004). *Book-Crossing Dataset*. [Online]. Available: http://www2.informatik.uni-freiburg.de/~cziegler/BX/

[6] F. Ortega, B. Zhu, J. Bobadilla, and A. Hernando, "Cf4j: Collaborative filtering for Java," *Knowl.-Based Syst.*, vol. 152, pp. 94–99, Jul. 2018.

[7] J. Bobadilla, F. Ortega, A. Hernando, and G. Glez-de-Rivera, "A similarity metric designed to speed up, using hardware, the recommender systems *k*-nearest neighbors algorithm," *Knowl.-Based Syst.*, vol. 51, pp. 27–34, Oct. 2013.

[8] D. Bokde, S. Girase, and D. Mukhopadhyay, "Matrix factorization model in collaborative filtering algorithms: A survey," *Procedia Comput. Sci.*, vol. 49, pp. 136–146, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050915007462

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

[10] 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.

[11] A. Abdelfattah, A. Haidar, S. Tomov, and J. Dongarra, "Fast Cholesky factorization on GPUs for batch and native modes in MAGMA," *J. Comput. Sci.*, vol. 20, pp. 85–93, May 2017.

[12] 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.

[13] S. K. Shinde and U. Kulkarni, "Hybrid personalized recommender system using centering-bunching based clustering algorithm," *Expert Syst. Appl.*, vol. 39, no. 1, pp. 1381–1387, Jan. 2012.

[14] J. C. Bezdek, W. Full, and R. Ehrlich, "FCM: The fuzzy c-means clustering algorithm," *Comput. Geosci.*, vol. 10, nos. 2–3, pp. 191–203, 1984.

[15] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms* (Advanced Applications in Pattern Recognition). New York, NY, USA: Springer, 2013.

[16] K.-L. Wu and M.-S. Yang, "Alternative c-means clustering algorithms," *Pattern Recognit.*, vol. 35, no. 10, pp. 2267–2278, 2002.

[17] H. Koohi and K. Kiani, "User based collaborative filtering using fuzzy C-means," *Measurement*, vol. 91, pp. 134–139, Sep. 2016.

[18] 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.

[19] C. Hai-Peng, S. Xuan-Jing, L. Ying-Da, and L. Jian-Wu, "A novel automatic fuzzy clustering algorithm based on soft partition and membership information," *Neurocomputing*, vol. 236, pp. 104–112, May 2017.
[20] C. Birtolo and D. Ronca, "Advances in clustering collaborative filtering by means of fuzzy C-means and trust," *Expert Syst. Appl.*, vol. 40, no. 17, pp. 6997–7009, 2013.
[21] R. Hu, W. Dou, and J. Liu, "ClubCF: A clustering-based collaborative filtering approach for big data application," *IEEE Trans. Emerg. Topics Comput.*, vol. 2, no. 3, pp. 302–313, Sep. 2014.
[22] M. Hasanzadeh-Mofrad and A. Rezvanian, "Learning automata clustering," *J. Comput. Sci.*, vol. 24, pp. 379–388, Jan. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877750317302247
[23] O. Fernando and B. Jesús, H. Antonio, and R. Fernando, "Using hierarchical graph maps to explain collaborative filtering recommendations," *Int. J. Intell. Syst.*, vol. 29, no. 5, pp. 462–477, 2014.
[24] M. Graham and J. Kennedy, "A survey of multiple tree visualisation," *Inf. Vis.*, vol. 9, no. 4, pp. 235–252, 2010.
[25] S. Kant and T. Mahara, "Nearest biclusters collaborative filtering framework with fusion," *J. Comput. Sci.*, vol. 25, pp. 204–212, Mar. 2018.
[26] F. Ortega, A. Hernando, J. Bobadilla, and J. H. Kang, "Recommending items to group of users using matrix factorization based collaborative filtering," *Inf. Sci.*, vol. 345, pp. 313–324, Jun. 2016.
[27] F. Ortega, R. Hurtado, J. Bobadilla, and R. Bojorque, "Recommendation to groups of users using the singularities concept," *IEEE Access*, vol. 6, pp. 39745–39761, 2018.
[28] M. Ge, C. Delgado-Battenfeld, and D. Jannach, "Beyond accuracy: Evaluating recommender systems by coverage and serendipity," in *Proc. 4th ACM Conf. Recommender Syst.*, New York, NY, USA, 2010, pp. 257–260.
[29] J. Kim and H. Park, "Sparse nonnegative matrix factorization for clustering," Georgia Inst. Technol., Atlanta, GA, USA, College Comput. Tech. Rep. 503, 2008. [Online]. Available: https://smartech.gatech.edu/handle/1853/20058
[30] J. Bobadilla and F. Serradilla, "The effect of sparsity on collaborative filtering metrics," in *Proc. 12th Australas. Conf. Australas. Database*. Darlinghurst, NSW, Australia: Australian Computer Society, vol. 92, 2009, pp. 9–18.
[31] C. C. Aggarwal and C. K. Reddy, *Data Clustering: Algorithms and Applications*. Boca Raton, FL, USA: CRC Press, 2013.
[32] X. Zhang, L. Zong, X. Liu, and J. Luo, "Constrained clustering with nonnegative matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 7, pp. 1514–1526, Jul. 2016.
[33] N. D. Buono and G. Pio, "Non-negative matrix tri-factorization for co-clustering: An analysis of the block matrix," *Inf. Sci.*, vol. 301, pp. 13–26, Apr. 2015.
[34] J. Yoo and S. Choi, "Weighted nonnegative matrix co-tri-factorization for collaborative prediction," in *Advances in Machine Learning*, Z.-H. Zhou and T. Washio, Eds. Berlin, Germany: Springer, 2009, pp. 396–411.
[35] C.-X. Zhang, Z.-K. Zhang, L. Yu, C. Liu, H. Liu, and X.-Y. Yan, "Information filtering via collaborative user clustering modeling," *Phys. A, Stat. Mech. Appl.*, vol. 396, pp. 195–203, Feb. 2014.
[36] X. Wang and J. Zhang, "Using incremental clustering technique in collaborative filtering data update," in *Proc. IEEE 15th Int. Conf. Inf. Reuse Integr. (IEEE IRI)*, Aug. 2014, pp. 420–427.
[37] H. Wu, Y.-J. Wang, Z. Wang, X.-L. Wang, and S.-Z. Du, "Two-phase collaborative filtering algorithm based on co-clustering," *J. Softw.*, vol. 21, no. 5, pp. 1042–1054, 2010.
[38] R. Gemulla, E. Nijkamp, P. J. Haas, and Y. Sismanis, "Large-scale matrix factorization with distributed stochastic gradient descent," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, 2011, pp. 69–77.
[39] R. Du, D. Kuang, B. Drake, and H. Park, "DC-NMF: Nonnegative matrix factorization based on divide-and-conquer for fast clustering and topic modeling," *J. Global Optim.*, vol. 68, pp. 777–798, Aug. 2017.
[40] C. Zhang, H. Wang, S. Yang, and Y. Gao, "Incremental nonnegative matrix factorization based on matrix sketching and k-means clustering," in *Intelligent Data Engineering and Automated Learning–IDEAL*, H. Yin *et al.*, Eds. Cham, Switzerland: Springer, 2016, pp. 426–435.
[41] R. Kannan, G. Ballard, and H. Park, "A high-performance parallel algorithm for nonnegative matrix factorization," *ACM SIGPLAN Notices*, vol. 51, pp. 9:1–9:11, Feb. 2016.
[42] K. Zeng, N. Wu, X. Yang, L. Wang, and K. K. Yen, "FHCC: A soft hierarchical clustering approach for collaborative filtering recommendation," *Int. J. Data Mining Knowl. Manage. Process*, vol. 6, no. 3, pp. 25–36, 2016.
[43] K. A. Vidhya and T. V. Geetha, "Rough set theory for document clustering: A review," *J. Intell. Fuzzy Syst.*, vol. 32, no. 3, pp. 2165–2185, 2017.
[44] Z. Deng, K.-S. Choi, Y. Jiang, J. Wang, and S. Wang, "A survey on soft subspace clustering," *Inf. Sci.*, vol. 348, pp. 84–106, Jun. 2016.
[45] R. Schachtner, G. Pöppel, A. M. Tomé, C. G. Puntonet, and E. W. Lang, "A new Bayesian approach to nonnegative matrix factorization: Uniqueness and model order selection," *Neurocomputing*, vol. 138, pp. 142–156, Aug. 2014.
[46] Q. Sun, J. Lu, Y. Wu, H. Qiao, X. Huang, and F. Hu, "Non-informative hierarchical Bayesian inference for non-negative matrix factorization," *Signal Process.*, vol. 108, pp. 309–321, Mar. 2015.
[47] B. Lakshminarayanan, G. Bouchard, and C. Archambeau, "Robust Bayesian matrix factorisation," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, vol. 15, G. Gordon, D. Dunson, and M. Dudík, Eds. Fort Lauderdale, FL, USA: PMLR, Apr. 2011, pp. 425–433.
[48] M. Luo, F. Nie, X. Chang, Y. Yang, A. G. Hauptmann, and Q. Zheng, "Probabilistic non-negative matrix factorization and its robust extensions for topic modeling," in *Proc. AAAI*, 2017, pp. 2308–2314.
[49] T. Virtanen, A. T. Cemgil, and S. Godsill, "Bayesian extensions to non-negative matrix factorisation for audio signal modelling," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar./Apr. 2008, pp. 1825–1828.
[50] K. Siminski, "Fuzzy weighted C-ordered means clustering algorithm," *Fuzzy Sets Syst.*, vol. 318, pp. 1–33, Jul. 2017.
[51] Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu, "Understanding of internal clustering validation measures," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2010, pp. 911–916.
[52] M. Muranishi, K. Honda, and A. Notsu, "Xie-beni-type fuzzy cluster validation in fuzzy co-clustering of documents and keywords," in *Soft Computing in Artificial Intelligence*, Y. I. Cho and E. T. Matson, Eds. Cham, Switzerland: Springer, 2014, pp. 29–38.
[53] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 8, pp. 841–847, Aug. 1991.
[54] G. Grekousis and H. Thomas, "Comparison of two fuzzy algorithms in geodemographic segmentation analysis: The fuzzy C-means and Gustafson–Kessel methods," *Appl. Geograph.*, vol. 34, pp. 125–136, May 2012.
[55] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Trans. Inf. Syst.*, vol. 22, no. 1, pp. 5–53, 2004.

REMIGIO HURTADO was born in 1989. He received the B.S. degree in systems engineering from Universidad Politécnica Salesiana, Ecuador, in 2012, the master's degree in information and software technology from the Instituto Tecnológico y de Estudios Superiores de Monterrey, México, in 2014, and the master's degree in computer science and technology from the Universidad Politécnica de Madrid, Spain, in 2017. He is currently a Lecturer with Universidad Politécnica Salesiana. He is a member of the research team in artificial intelligence and assistive technology, in collaboration with the Universidad Politécnica de Madrid. His research interests include the recommender systems and natural language processing.

JESÚS BOBADILLA received the B.S. degree in computer science from the Universidad Politécnica de Madrid and the Ph.D. degree in computer science from Universidad Carlos III. He is currently an Associate Professor with the Department of Intelligent Systems, Universidad Politécnica de Madrid. He is an in-charge of the filmaffinity.com Research Team, where he focuses on the collaborative filtering kernel of the website. He was also a Researcher with the International Computer Science Institute, Berkeley University, and also with Sheffield University. He is the Head of the research group. He is a Habitual Author of programming languages books working with McGraw-Hill, Ra-Ma, and Alfa Omega publishers. His research interests include information retrieval, recommender systems, and speech processing.

RODOLFO BOJORQUE was born in 1982. He received the B.S. degree in systems engineering from Universidad Politécnica Salesiana, Ecuador, in 2006, the master's degree in security information and communications technologies from the Universidad Autónoma de Barcelona, Spain, in 2015, and the master's degree in computer science and technology from the Universidad Politécnica de Madrid, Spain, in 2017.

He is currently a Lecturer with Universidad Politécnica Salesiana. He is also a member of the research team in artificial intelligence and assistive technology, in collaboration with the Universidad Politécnica de Madrid. His research interests include the machine learning techniques and intelligent systems.

FERNANDO ORTEGA was born in 1988. He received the B.S. degree in software engineering and the Ph.D. degree in computer sciences from the Universidad Politécnica de Madrid, Madrid, in 2010 and 2015, respectively. He is currently an Assistant Professor with the Department of Intelligent Systems, Universidad Politécnica de Madrid. He has published several papers in the most relevant international journals. He also actively collaborates in various projects with technology companies. His main research fields include machine learning, data analysis, and artificial intelligence.

XIN LI received the B.Sc. and M.Sc. degrees from Jilin University, China, and the Ph.D. degree from Hong Kong Baptist University. She was a Postdoctoral Teaching Fellow with Hong Kong Baptist University. In 2010, she joined the Beijing Institute of Technology, where she is currently an Associate Professor with the School of Computer Science. She has published more than 30 papers in refereed international journals and conference proceedings, including AAAI, IJCAI, ICML, KAIS, TKDE, and TOC. Her research interests include representation learning and deep reinforcement learning, with its applications to healthcare and robotics. Her team received the Best Paper Award at IPCCC 2013. She also served/serves as a PC for AI conferences including CIKM, WSDM, and IJCAI. As a Principle Supervisor, she has been supervising multiple projects funded by the National Natural Science Foundation of China, National Program on Key Basic Research Foundation, and Doctoral Programs Foundation of Ministry of Education, China, since 2011.