UCLouvain

LOUVAIN
School of Management

Louvain School of Management

Including fairness in recommendation systems by decorrelation with sensitive variables

Analysis of its impact on the factorization error

Auteur : Modeste Bodehou

Promoteur : Marco Saerens

Année académique 2021 - 2022

Travail de fin d'études (TFE) en vue d'obtenir le titre

de Master (60) en Sciences de Gestion

Horaire de jour

Abstract

Recommendation systems are filtering algorithms developed to predict the preference of a user for a given item. Those systems may be categorized into two groups. In the first group, i.e. collaborative filtering, it is assumed that people who agreed in the past will also agree in the future. Comparing the past preferences of the users then allows one to estimate their future preferences. The second approach, content based filtering, relies on the expansion of each item into a set of characteristics. The basic assumption is that a user who prefers a given item should also prefer items with similar characteristics. In this thesis, the focus is put on the first approach, i.e. collaborative based filtering because this class of methods has the advantage of not requiring an understanding of the items characteristics. However, although collaborating filtering has proven to be able to provide good performance in term of prediction error, those algorithms tend to reproduce in the predictions possible correlations/discriminations (with respect to some sensitive variables) existing in the training data set. This characteristic may cause in certain circumstances, ethical issues. Therefore, it is in some cases preferable to make the predictions, independent with respect to some sensitive variables so as to introduce more fairness in the recommendation. This additional constraint may (or not) be achieved at the expense of a worse prediction error as compared to the solution with discriminations. The goals of this thesis are twofold. First, two main collaborative filtering algorithms (matrix factorization based method, and the K-nearest neighbors algorithm) are reviewed and compared against each other. Second, fairness constraint is introduced, and its impact on the prediction error is closely analyzed. Ability of the developed algorithm to reduce possible discriminations is demonstrated through various discrimination measures, namely the covariance, and correlation.

Contents

List of Figures

1

Introduction

1.1 Application of data analytics in management

The rapid growing of the e-commerce offers an unprecedented access to a large amount of users data [1]. Being able to analyze those data so as to extract meaningful information is of strategic interest in almost all modern business activities. For example, a company that offers a vast variety of products can significantly improve its profit if it is able to predict the users interests and consequently adapt its recommendations to each user. Understanding the users preferences is paramount to the marketing analysis since it enables a better segmentation of the market, a better estimation of the attractiveness of each market segment, and a better customer relationship management [2]. However, the advantage of data analytics goes beyond marketing applications. A logistic company can use data analytics for operations management. Given the predictions of the customers interests, an optimized way to route the goods to different warehouses and sales center can be found [3]. The manufacturing direction of a company can adapt its process according to the predicted demands so as to minimize the unit cost. Data analytics is also useful for the health sector to predict for example the side effects of some drugs [2], for the police to predict potential unsecure places [4], for financial places to predict the evolution of a stock [5], for wireless communication systems to design cognitive radios that smartly sense their environment [6], [7], for geospatial intelligence [8], computer security [9], etc. All these examples show that big data offers a great support to decision making for companies management. Companies that will be able to extract non-trivial but useful patterns from the data will definitely have a strategic advantage on the market because they will better match the users demands and be able to optimize their processes.

1.2 Recommendation systems

The rapid development of companies like Amazone, Google, Netflix, undoubtedly fosters the development of recommendation systems [10], [11]. Assuming that each user has previously rated a set of items, the fundamental question is the following: is it possible to predict the score each user would give for the items he has not rated? The capability to properly answer this question is at the core of the business strategy of many companies and is nowadays an important issue in business analytics. Two approaches are commonly used to address this problem and rely on different assumptions. The first approach is collaborative filtering [12]. It assumes that the agreement level between people does not change over the time. That means, based on the evaluation of the level of agreement between a given user and other ones, it should be possible to predict the score that a user of interest will give for a given item he has not rated, without knowing the characteristics of that item. The second approach is content based filtering. There, the recommendation is based on the expansion of each item into a set of a predefined characteristics [13]. Then items with similar characteristics is susceptible to be rated similarly by the user. One of the advantages of the content based filtering is that it requires less data to start because it relies mainly on the profile description of the user with respect to some characteristics of the items. In this thesis, it is assumed that we have enough data to evaluate similarities between users so as to rely on collaborative filtering. Collaborative filtering is nowadays well used in the e-commerce by several companies (Amazone, Netflix, Youtube, etc), in the social media (e.g. Facebook, Twitter, etc), to recommend products/items that the user is susceptible to like the most and consequently provides to those companies a significant competitive advantage. The usage of recommendation systems also supposes that the user profile is well defined. Indeed, when a computer or an account is shared by several users, the recommendation may be less suitable to the active member [14]. This issue is outside the scope of the present work.

1.3 Fairness in management

Fairness is essential in companies management. A responsible manager should bring value not only to his company or his industry, but also to the entire society [15]. Recent literature has proven that industries that do not integrate ethics in their strategy and goals, cannot thrive in the long term [16]. However, the basic assumption behind collaborative filtering may sometimes lead to some ethical issues [17], [18]. Indeed, since the past level of agreement is assumed to be preserved, initial patterns in the data set are also susceptible to be preserved in the predictions, and as such, may lead to some discriminations. An example of discrimination has been provided in [19]. In the 1930s, some banks in the united-states used to rely on some historical data of neighborhoods to evaluate the risk of individuals wishing to contract a loan. Although the model is not a priori based on the race, it has been observed that there

is a correlation between the predicted risk and the race. Indeed, black people appear to be in general more risky than white people: there is a discrimination based on the race.
Another discrimination, now based on gender, has been reported in [18]. A system has been developed to recommend studies to young people, based on historical data. However, since in the historical data, there are relatively few women in engineering and sciences as compared to men, the system tends to avoid recommending scientific studies to women, leading to worse recommendation for women. Such a discrimination may have a bad impact on women's careers. In marketing, bias/discrimination in recommendation systems can also make some products a priori less suited for women/men although a woman/man may be well interested in those products [20]. Discrimination can also occur between ethnicities or other kind of people groups in healthcare, human resources management, or for fraud detection [21]. This none-exhaustive list highlights the importance of introducing fairness or ethics in recommendation systems, thus avoiding reproducing in the predictions, some patterns or correlations pre-existing in the database. Integrating fairness in recommendation systems will help managers to take more responsible decisions, will reduce discrimination and inequalities, provide more justice in the society, in line with the Sustainable Development Goals (SDG) [22].

1.4 Goal of the thesis

This thesis aims at analyzing some of the well known collaborative filtering algorithms, namely the K-nearest neighbors (KNN) method [23], and the factorization based method [24]. The KNN approach is more intuitive and relies on the concept of similarity between users. Matrix factorization tends to construct a feature matrix for the users and the items, respectively. The analysis of the two methods will be carried out on realistic data downloaded from movielens, allowing to compare their performance. In this work, the performance quantifies the accuracy of the prediction. It will be defined as a root mean square error (RMSE) or as a mean absolute error (MAE) between the prediction and the ground truth. To do so, the available data set will be divided into two different parts, namely the training data and the validation data. The training data serves to run the algorithms and the validation data is used to estimate the accuracy of the precision. Once the two methods have been compared with each other, fairness will be introduced in the collaborative filtering. The impact of that introduction in the performance of the algorithm will be closely analyzed. A parametric study will be carried out to show the main features affecting the algorithm performance.

2

Collaborative filtering

This chapter provides a theoretical review of the collaborative filtering algorithms studied in this thesis, namely the K-Nearest Neighbor (KNN) method [23] and the matrix factorization based method [24].

Let us assume that we have m items (objects) that can be scored by n users. The rating is ranked from r_{min} to r_{max} depending on the interest of the user to the considered item. One can define a matrix V of size $n \times m$ containing the score of each user for a given item. In practice, V is not completely known because each user only scored a few items. The goal of collaborative filtering consists of filling V, i.e. estimating the missing elements on the basis of the existing elements. In this way, one can recommend to the users, item that they are most likely to appreciate. We define an indicator matrix $I \in \{0,1\}^{n \times m}$, such that $I_{ij} = 1$ if the user u_i has scored the object o_j. Given that each user only scores a few items, I is in practice very sparse.

2.1 The K-Nearest Neighbors (KNN) method

The KNN is the most intuitive collaborative filtering method. This method can be formulated as a user based or as an item based. In this work, we will focus on the user based formulation. Let us assume that it is required to predict the score of an item/object o_j, for a given user u_i. The KNN idea consists of finding the K-closest (or similar) users to the user of interest, that have rated o_j. Once those K-closest users are found, the prediction is computed as a weighted average of the scores provided by the K-nearest neighbors.

Different similarity measures can be used, among which the Pearson correlation coefficient, the Jaccard index, etc [25]. A comparison between those similarity measures for the KNN method has been studied by [26]. In the present work, we have used the Pearson correlation coefficient. The Pearson coefficient of similarity

between the users u_i and u_j is defined as:

$$sim(u_i, u_k) = \frac{\sum_{j=1}^{m} I_{ij} I_{kj} (V_{ij} - \overline{V}_i)(V_{kj} - \overline{V}_k)}{\sqrt{\sum_j^m I_{ij} I_{kj} (V_{ij} - \overline{V}_i)^2} \sqrt{\sum_j^m I_{ij} I_{kj} (V_{kj} - \overline{V}_k)^2}}, \tag{2.1}$$

where the summation is carried out over all the items o_j; $\overline{V}_i$ is the average score provided by the user u_i, i.e.

$$\overline{V}_i = \frac{\sum_{j=1}^{m} I_{ij} V_{ij}}{\sum_{j=1}^{m} I_{ij}}, \tag{2.2}$$

and $\overline{V}_k$ is the average score provided by the user u_k. The Pearson similarity can be geometrically interpreted as the cosine of the angle between the user vector V_i and the user vector V_k, from which their average scores have been subtracted.

The K nearest neighbors are then selected as the K users with the highest similarities with the user of interest u_i. Once those neighbors are selected, the prediction P_{ij} is obtained as:

$$P_{ij} = \frac{\sum_{k \in K} sim(u_i, u_k) I_{kj} V_{kj}}{\sum_{k \in K} sim(u_i, u_k) I_{kj}}, \tag{2.3}$$

where the set $k \in K$ corresponds to the K nearest neighbors.

2.2 Matrix factorization based methods

The matrix factorization methods tend to approximate the score matrix V of size $n \times m$ with the product or two matrices of lower ranks U and M [27], i.e.

$$V \approx P = g(U^T M), \tag{2.4}$$

with U a user matrix of size $f \times n$, and M an item matrix of size $f \times m$. Those methods have played an important role in the progress of the Netflix prize. Many techniques have been proposed to solve the matrix factorization problem, among which the alternating least squares (ALS) method [28], the multiplicative update method [29], the stochastic gradient methods [30], etc. In this work, we have used the approach proposed in [27]. The matrices U and M are obtained after minimizing the square error between V and P for the rated values in V. The cost function can therefore be defined as:

$$e = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{M} I_{ij} (V_{ij} - g(U_i, M_j))^2 + \frac{k_u}{2} \sum_{i=1}^{N} ||U_i||^2 + \frac{k_m}{2} \sum_{j=1}^{M} ||M_j||^2, \tag{2.5}$$

where the coefficients k_u and k_m serves as a regularization factor to prevent overfitting. The function g in (2.4), and (2.5) is simply used to bound the predicted values in

a given interval or to add offset variables. For example, assuming that the scores belong to the interval $[r_{min},\ r_{max}]$, the function g can be defined as:

$$\begin{aligned} g(U_i, M_j) &= r_{min} \quad \text{if} \quad U_i^T M_j \leq r_{min} \\ g(U_i, M_j) &= U_i^T M_j \quad \text{if} \quad r_{min} \leq U_i^T M_j \leq r_{max} \\ g(U_i, M_j) &= r_{max} \quad \text{if} \quad U_i^T M_j \geq r_{max} \end{aligned}$$

The derivative of (2.5) with respect to U_i and M_j is then obtained as:

$$-\frac{\partial e}{\partial U_i} = \sum_{j=1}^{M} I_{ij}((V_{ij} - g(U_i, M_j))M_j) - k_u U_i,\ \ i = 1...n \tag{2.7}$$

$$-\frac{\partial e}{\partial M_j} = \sum_{i=1}^{N} I_{ij}((V_{ij} - g(U_i, M_j))U_i) - k_m M_j,\ \ j = 1...m \tag{2.8}$$

From there, the gradient descent algorithm can be performed. At each iteration, the matrices U and M are updated as:

$$U \leftarrow U - \mu\frac{\partial e}{\partial U} \tag{2.9}$$

$$M \leftarrow M - \mu\frac{\partial e}{\partial M} \tag{2.10}$$

The coefficient μ is the learning rate. μ should not be too small to avoid high number of iterations for convergence. However, high values of μ can make the algorithm diverge. So, in practice, a compromise should be found for the value of the learning rate. It is also clear that the initial/starting point of U and M will also affect the performance of the algorithm. The authors of [27] propose a starting point randomly distribute around the average of the rating matrix, i.e.

$$U_{ij}, M_{ij} = \sqrt{\frac{\overline{V} - r_{min}}{f}} + n(r), \tag{2.11}$$

where $n(r)$ is a random noise with uniform probability density between $-r$ and r. $\overline{V}$ is the average of the rating matrix, defined as:

$$\overline{V} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} I_{ij}V_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{m} I_{ij}}. \tag{2.12}$$

2.3 Improvement of the method

In some cases, adding an offset to the factorization, may provide better results. That means, the prediction function can be defined as [31]:

$$P_{ij} = U_i^T M_j + \alpha_i + \beta_j, \tag{2.13}$$

where α_i and β_j are the unknown users and items offsets, respectively. Those offsets should be optimized simultaneously with the factorization matrices U and M. The cost function is now written as:

$$e = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}(V_{ij} - g(U_i, M_j, \alpha_i, \beta_j))^2 + \frac{k_u}{2}\sum_{i=1}^{N}||U_i||^2 + \frac{k_m}{2}\sum_{j=1}^{M}||M_j||^2 + \frac{k_b}{2}(\sum_{i=1}^{N}\alpha_i^2 + \sum_{j=1}^{M}\beta_j^2), \tag{2.14}$$

The derivatives of the cost function with respect to α_i and β_j are then given by:

$$-\frac{\partial e}{\partial \alpha_i} = \sum_{j=1}^{M} I_{ij}((V_{ij} - g(U_i, M_j, \alpha_i, \beta_j))M_j) - k_b\alpha_i, \;\; i = 1...n \tag{2.15}$$

$$-\frac{\partial e}{\partial \beta_j} = \sum_{i=1}^{N} I_{ij}((v_{ij} - g(U_i, M_j, \alpha_i, \beta_j))U_i) - k_b\beta_j, \;\; j = 1...m \tag{2.16}$$

and are used to compute the gradient required to optimize α_i and β_j.

2.4 Performance of the recommendation

The performance of the recommendation can be measured by comparing the prediction of the filtering algorithm to the ground-truth. However in practice, the actual values of the predicted scores are not known. For validation purpose, the training and validation data sets are randomly chosen in the rating matrix. This technique is used for the Netflix prize [32].

Let us define an indicator matrix $J \in \{0, 1\}^{n\times m}$, for the training data. That means, $J_{ij} = 1$ if V_{ij} belongs to the training data, and 0 otherwise. One should note that, since the training data is selected from the available scores, $J_{ij} = 1 \Rightarrow I_{ij} = 1$. The validation data corresponds to the data for which $I_{ij} = 1$ and $J_{ij} = 0$. It is clear that the accuracy of the prediction will depend on the selection procedure of the training and validation data. This aspect will be evaluate in the next chapter. In the present work, the training data selection has been inspired by the one used for the Netflix prize [32]. For each user, we have randomly selected in the available data, a fix number of items, serving for validation purpose. To do so, we have generated for each user, a vector of uniformly distributed random variables between 0 and 1. This set of vectors forms a new matrix J_a of size $(n \times m)$, where each line corresponds to a given random user vector. Each element of J_a is then multiplied by its corresponding element in the indicator matrix I, leading to the matrix J_2, i.e.

$$J_2 = J_1 \odot I, \tag{2.17}$$

where the symbol $\odot$ represents a point by point multiplication. For each user (row), the validation data index are selected as the ones corresponding to the highest values

in the corresponding row of J_2. The training data set corresponds to the available scores, which have not been selected for the validation. Once the validation and training data selected, the accuracy of the prediction can be computed as the root mean square error:

$$RMSE(P,V) = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{m} W_{ij}(V_{ij}-P_{ij})^2}{\sum_{i=1}^{n}\sum_{j=1}^{m} W_{ij}}}, \tag{2.18}$$

or as the mean absolute error:

$$MAE(P,V) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} W_{ij}|V_{ij}-P_{ij}|}{\sum_{i=1}^{n}\sum_{j=1}^{m} W_{ij}}, \tag{2.19}$$

where W is the indicator matrix for the validation data.

2.5 Including fairness in the recommendation

2.5.1 Fairness modeling

The previous section has addressed the factorization based recommendation algorithm. As explained in the introduction, such a recommendation may not be "fair" with respect to some groups of people. The notion of fairness is somehow subjective because it is intimately linked to ethic, social values. The recommendation algorithm can be considered as being fair if its output is independent with respect to some sensitive variables. The sensitive variables depend on the context, and can be for example: the sex, the race, the ethnic group, the religion, etc. Since the prediction is made from the available data set, a bias in the available data may appear in the prediction. It is therefore important to impose in the algorithm, the absence of those sensitive bias, so as to make the recommendation more fair.

In probability theory, a random variable X (with probability density function $f_x(x)$) is independent of a random variable Y, if and only if:

$$f_x(x|y) = f_x(x) \tag{2.20}$$

This definition can be restricted to different set of groups G_x and G_y [33]. In that case, one have:

$$f_x(x \in G_x | y \in G_y) = f_x(x \in G_x). \tag{2.21}$$

A simple way to quantify fairness, considering two groups G_x and G_y, may consist of imposing that:

$$\sum_{j=1}^{m} |E_j(x \in G_x) - E_j(y \in G_y))| < \epsilon, \tag{2.22}$$

where ϵ is a very small number, $E_j(x \in G_x)$, $E_j(y \in G_y)$ are the average of the prediction for item j and group G_x, G_y respectively. That means, for a given item, it is

required that the average of the prediction for the two groups should be approximately the same. Considering that in the initial data set, the two groups may not have the same average, it is proposed in [18], [34] to impose fairness as

$$\sum_{j=1}^{m} |E_j(x \in G_x) - V_{jx}| - |E_j(y \in G_y) - V_{jy}|| < \epsilon, \tag{2.23}$$

where V_{jx}, V_{jy} are the average score in the initial data set for item j and group G_x, G_y, respectively.

Another way to define fairness consists of using the covariance [35]. Considering two vectors x and y, the covariance between x and y is defined as:

$$cov(x, y) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) \tag{2.24}$$

The covariance quantifies the joint deviation of the variables x and y with respect to their respective average. It is a way to quantify the correlation between the variables of interest. The correlation $cor(x, y)$ between the variables x and y [36], can be obtained from the covariance by simple renormalization:

$$cor(x, y) = \frac{cov(x, y)}{\sigma_x \sigma_y}, \tag{2.25}$$

where σ_x and σ_y are the standard deviation of x and y, respectively. This normalization enables the correlation to stay in the interval [-1, 1]. In this work, the covariance will be used as a model for the fairness. The covariance defined in (2.25) can be rewritten as [35]:

$$cov(x, y) = \frac{1}{n-1} x^T H y, \tag{2.26}$$

where $H = I - \frac{1}{n} e e^T$ is a matrix of size $n \times n$, I is the identity matrix of size $n \times n$, and e is a vector of ones, and size n.

Other fairness metrics have been proposed in the literature, for example based on parametric statistics test [37].

2.5.2 Fairness optimization

Including fairness in the recommendation requires the joint optimization of the prediction error and the covariance, modeling the fairness. Considering for example one sensitive variable s_1, and a factorization without offset variables, the cost function to be minimized can then be defined as:

$$\begin{aligned} e = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{M} I_{ij} (V_{ij} - U_i^T M_j)^2 + \frac{k_u}{2} \sum_{i=1}^{N} ||U_i||^2 + \frac{k_m}{2} \sum_{j=1}^{M} ||M_j||^2 \\ + \frac{\lambda}{2} \sum_{j=1}^{m} cov(U^T M_j, s_1)^2, \end{aligned} \tag{2.27}$$

Inserting, expression (2.26), in (2.27), leads to:

$$e = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}(V_{ij} - U_i^T M_j)^2 + \frac{k_u}{2}\sum_{i=1}^{N}||U_i||^2 + \frac{k_m}{2}\sum_{j=1}^{M}||M_j||^2 + \frac{\lambda}{2}\sum_{j=1}^{m}(s_1^T H U^T M_j)^2, \tag{2.28}$$

which we have been rewritten as:

$$e = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}(V_{ij} - U_i^T M_j)^2 + \frac{k_u}{2}\sum_{i=1}^{N}||U_i||^2 + \frac{k_m}{2}\sum_{j=1}^{M}||M_j||^2 + \frac{\lambda}{2}\sum_{j=1}^{m}(Q^T U^T M_j)^2, \tag{2.29}$$

where Q is a vector of size n. The derivatives of the covariance term can then be computed as [35]:

$$-\frac{\lambda}{2}\frac{\partial}{\partial U_i}\left[\sum_{j=1}^{M} cov(U^T M_j, s_1)^2\right] = -\lambda\sum_{j=1}^{M}[(Q_i M_j).(Q^T U^T M_j)], \tag{2.30}$$

$$-\frac{\lambda}{2}\frac{\partial}{\partial M_j}\left[\sum_{j=1}^{M} cov(U^T M_j, s_1)^2\right] = -\lambda[(Q^T U^T M_j).(Q^T U^T)] \tag{2.31}$$

Those derivatives are added to the expressions (2.7) and (2.8) to compute the gradient of the cost function.

Extension to the case of several sensitive variables is straightforward. One simply needs to add other covariance terms.

3

Results

This chapter provides numerical results for the recommendation algorithms developed in the previous chapter. The first objective consists of comparing the performance of the KNN, and that of the factorization based method. After that, fairness will be included in the objective function as described in section 2.5.2. The impact of the fairness constraint on the prediction accuracy will be studied. Finally, the fairness of the recommendation, modeled as the correlation (or the covariance) with the sensitive variables will be analyzed.

3.1 Database

We have used a data set collected by the GroupLens Research Project of the University of Minnesota [38]. The group has collected the data set on the movielens web site (movielens.umn.edu) between 1997 and 1998. The data set contains scores provided by users on various movies, as well as some demographic information of the users (age, gender, occupation, etc). Users with incomplete data (less than 20 ratings) or not complete demographic information have been removed. The resulting database contains 100000 ratings from 943 users and 1682 items (movies), and the demographic information of each user. The ratings vary from 1 to 5. The data set has been downloaded at movielens.

3.2 Recommendation based on KNN

The KNN is applied on the data set to predict 10 scores, randomly chosen for each user, following the procedure described in section 2.4. Note that the data set has already been processed so as to garantee at least 20 ratings per user. So it is possible to randomly chose 10 ratings per user for validation. The accuracy of the prediction is evaluated with the Root Mean Square Error (RMSE), and the Mean Average error (MAE), as defined in equations (2.19) and (2.18). The KNN has been run for different numbers of nearest neighbors K. Figure. 3.1 shows the evolution of

the MAE and that of the RMSE as a function of K. It can be seen that the error

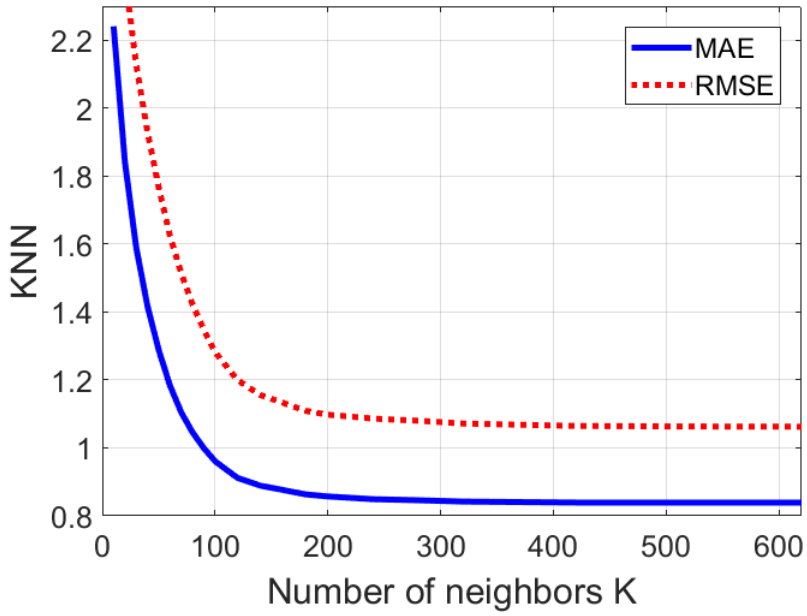


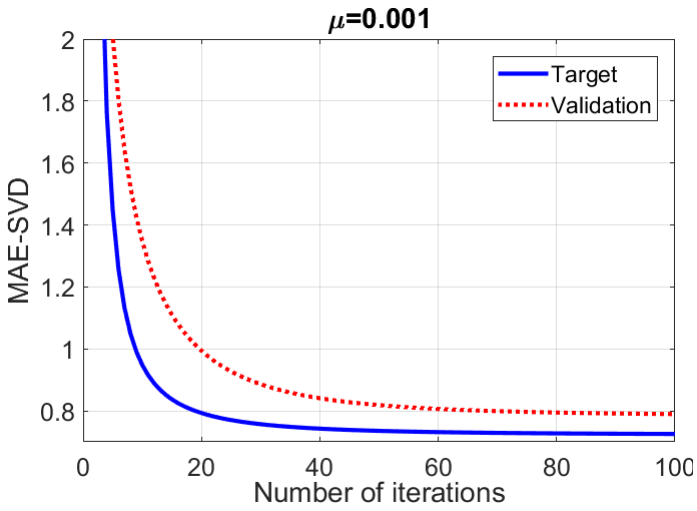
Figure 3.1: Prediction error of the KNN as a function of the number of considered neighbors.

decreases, when considering a higher number of neighbors until the minimum value $MAE = 0.84$ and $RMSE = 1.06$.

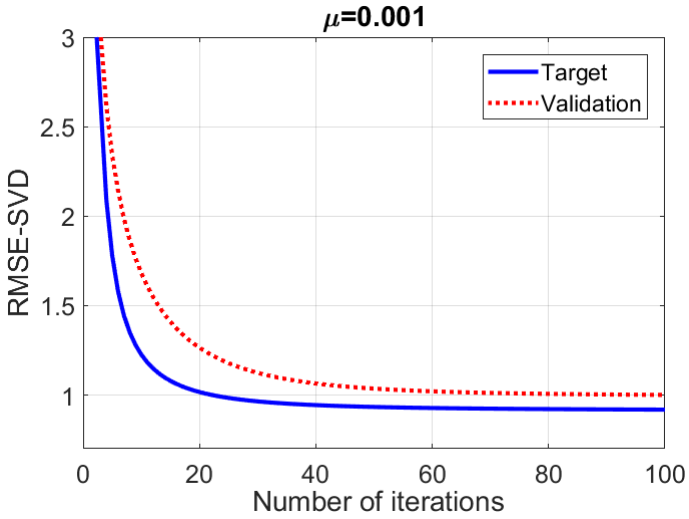
3.3 Recommendation based on matrix factorization

3.3.1 Factorization without offset variables

This section uses the matrix factorization technique described in the previous chapter to predict 10 ratings randomly chosen per user. First, the factorization is made without offset variables (i.e. $\alpha_i = \beta_j = 0$). The prediction is then given by: $P = U^T M$, with U, and M being the factorization matrices of size $f \times n$, and $f \times m$ respectively. We have fixed $f = 10$, and $\mu = 0.001$. The MAE and the RMSE are shown as a function of the iteration number in Figure 3.2. In those figures, the MAE and the RMSE have been computed on the training/target data set (i.e the data set one which the factorization has been made), and on the validation data set. Of course, since the validation data is different from the training data, one should expect a degradation of the performance. Anyway a good convergence can be observed. The optimal MAE, and RMSE on the training data are $MAE_t = 0.72$, and $RMSE_t = 0.92$. Applying now the factorization on the validation data, we obtained



(a)



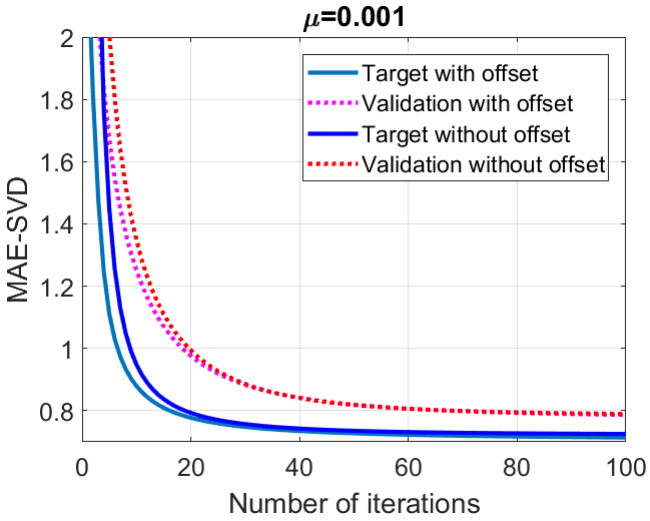
(b)

Figure 3.2: Prediction error of the factorization based method, as a function of the number of considered neighbors.

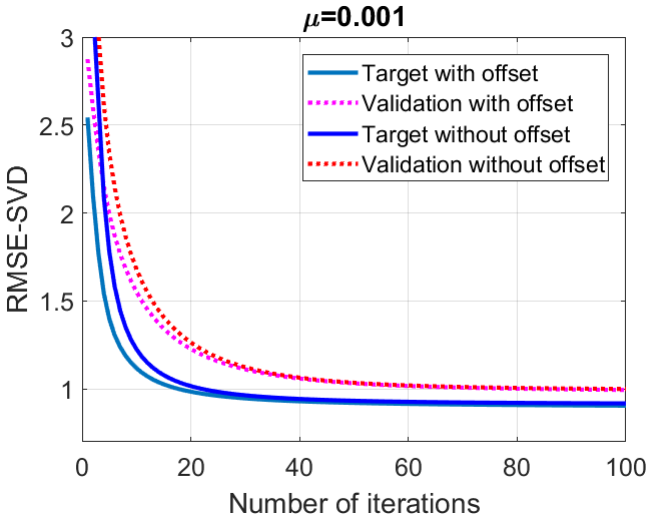
$MAE_v = 0.79$, and $RMSE_v = 1.001$. Given that with the KNN, we have obtained $MAE_v = 0.84$, and $RMSE_v = 1.06$, it can be concluded that the factorization based method provides better results than the KNN. In the rest of the thesis, we will then use the factorization based method.

3.3.2 Factorization with offset variables

The previous analysis was carried out without the usage of offset variables α_i, β_j (see equation (2.13)). Now, we have added user offset α_i. The offset is optimized simultaneously with the factorization matrices. Figure 3.3 shows the results. We



(a)



(b)

Figure 3.3: Prediction error of the factorization based method as function of the number of iterations: impact of adding offset variables.

have observed a very small improvement ($MAE_t = 0.725$, $MAE_v = 0.789$ without

offset variables, and $MAE_t = 0.714$, $MAE_v = 0.786$ with offset variables). In the rest of the chapter, we will then consider the matrix factorization method with user offset variables.

3.3.3 Influence of the parameter f on the performance

In the previous example, we have chosen $f = 10$. A legitimate question is: what about other choice of f? This analysis has been reported in Figure 3.4. As clearly

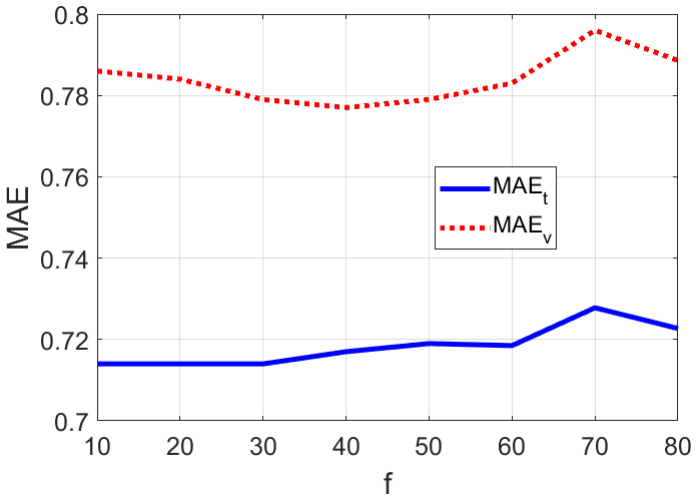


Figure 3.4: Impact of the factorization matrix size, on the prediction error.

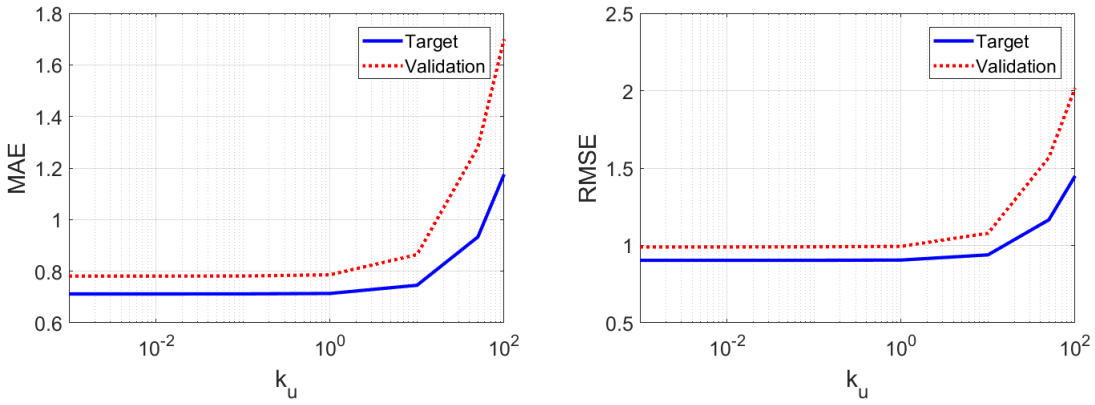
seen, the value of f has not a significant impact on the performance. Therefore, we will choose $f = 10$ in the rest of the analysis.

3.3.4 Influence of the regularization coefficient

This section studies the impact of the regularization term on the prediction error. Figure 3.5 illustrates the prediction error as a function of the regularization coefficient k_u, assuming that $k_u = k_m$. One can conclude that the regularization term has almost no impact up to $k_u = k_m = 1$. Above that, the performance starts deteriorating. In the rest of the work, we have fixed $k_u = k_m = 0.5$.

3.3.5 Impact of the size of the validation data set

It is clear that the statistics of the chosen validation data set will impact the prediction error. This section analyzes the impact of the relative size of the validation data with respect to the training data. For that purpose, we run the optimization considering different validation data set (number of validation data per user). The result is shown in Figure 3.6. One can conclude that the prediction error increases for higher values



(a) (b)

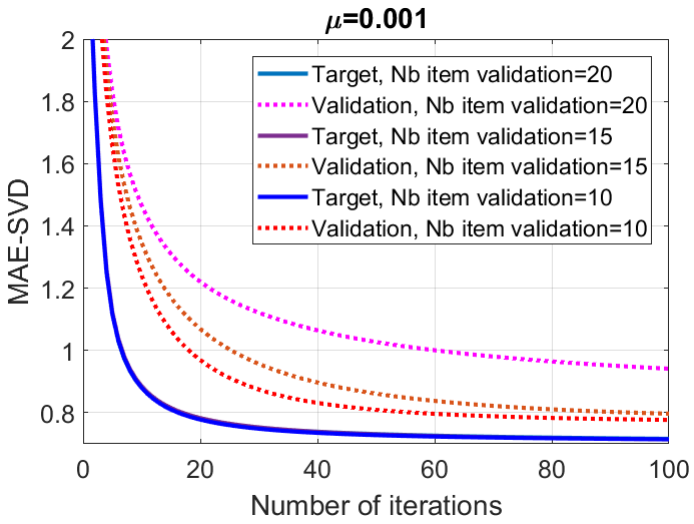
Figure 3.5: Impact of the regularization coefficient, on the averaged prediction error.

of the number of validation items. However, the error computed on the training data does almost not change.

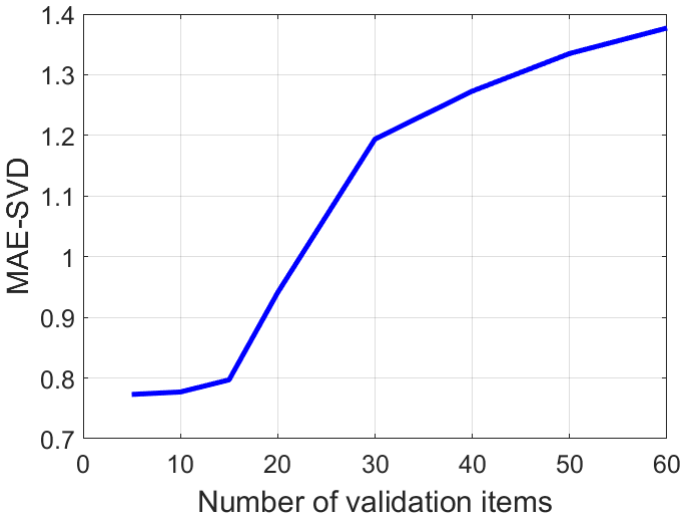
3.4 Fair recommendation based on matrix factorization

Up to now, the recommendation has been made without including fairness. This section analyzes the impact of including fairness. We have chosen the sensitive variables $s_1 = sex$ and $s_2 = age$. The sex is modeled as a binary variable, whose value equal one for men, and zero for women. The covariance of each sensitive variable with the predicted scores is minimized as explained in the previous chapter. We have fixed for the sex, $\lambda_1 = 2.35$, and for the age, $\lambda_2 = 0.00235$. Figure 3.7 compares the fair prediction error with the unfair prediction error. As can be seen, including fairness does almost not change the prediction accuracy. Let's now see the impact of including fairness on the covariance (or the correlation) with the sensitive variables. For this purpose, we have computed for each item, the covariance/correlation between the prediction, and a given sensitive variable; then we take the average (over all the items) of the absolute value of the computed covariance/correlation. Figures 3.8 and 3.9 show the results for the sensitive variables "sex" and "age" respectively.

It is clear that the algorithm has significantly included fairness, as compared to the case without fairness. This fairness benefit has been achieved without loosing prediction error performance. This result can be explained by the non-convexity property of the optimization problem, leading to several local minimums. Moreover, it is interesting to see from Figure 3.8 that the correlation oscillates as a function of the iteration number when fairness is not included. So, depending on the number of



(a)

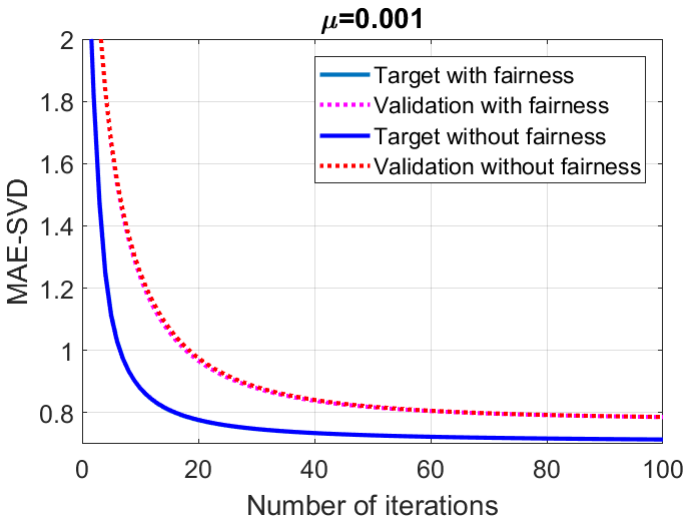


(b)

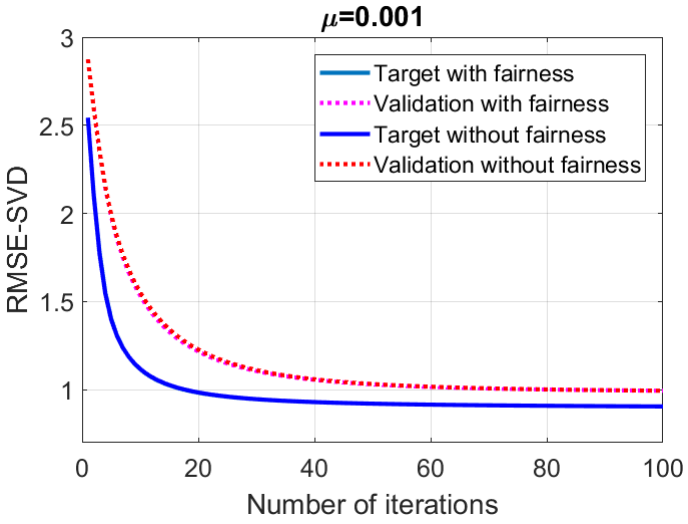
Figure 3.6: Impact of the size of the training data relative to the validation data, on the average prediction error.

iterations at which the optimization has been stopped, the correlation preexisting in the initial data set may be amplified or not.

In the next step, we analyze the impact of varying λ_1 and λ_2, on the covariance and the correlation. For that purpose, the optimization is performed for different values of (λ_1, λ_2), with $\lambda_2 = 0.001\lambda_1$. The variation of the average covariance/correlation



(a)

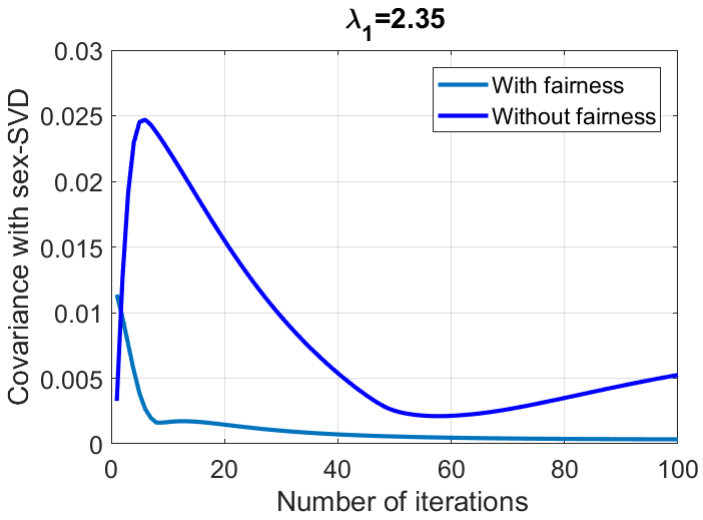


(b)

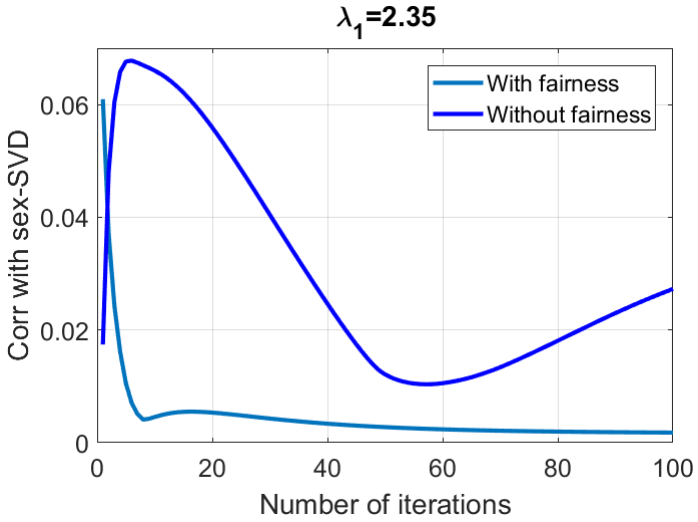
Figure 3.7: Convergence of the prediction error, with and without including fairness.

as a function of λ_1 for the sex, and λ_2 for the age, is shown in Figure 3.10.

It can be seen that, increasing the value of λ improves the performance. It makes sense because λ determines the weight put on the fairness term. However, the variation is non linear. The curves are convex, which means that the variation of the gain in fairness, for higher values of λ is smaller. However, increasing too much the value of λ without decreasing the learning rate μ can make the solution diverge. This



(a)

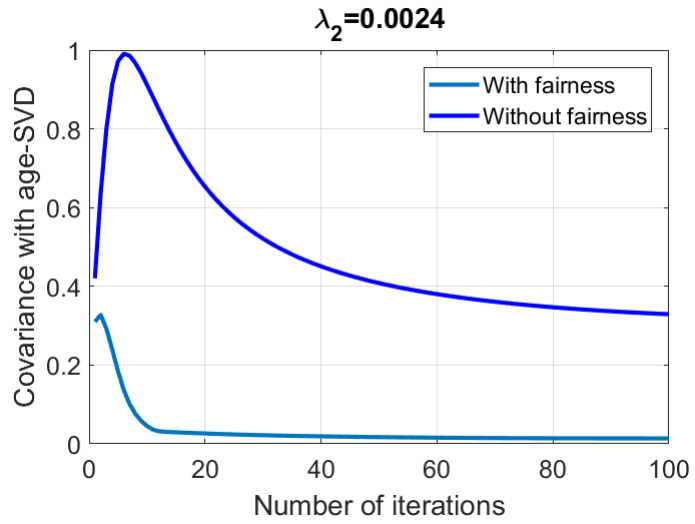


(b)

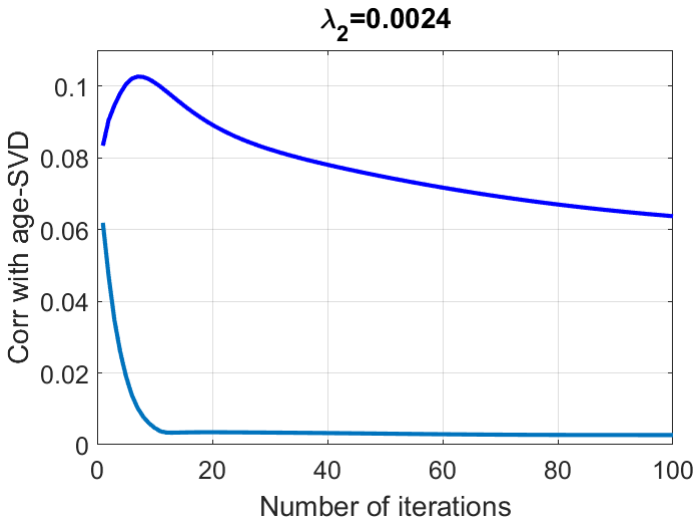
Figure 3.8: Illustration of the convergence of the covariance/correlation with the sex, when fairness is included. The solution without fairness is compared to the solution with fairness.

problem is illustrated in Figure 3.11 and 3.12. Figure 3.11 shows the convergence of the MAE and the RMSE for different values of λ. One can observe some pics, which is explained by the locally too high value of the learning rate.

This problem is more drastic when analyzing the convergence of the covari-



(a)



(b)

Figure 3.9: Illustration of the convergence of the covariance/correlation with the age, when fairness is included. The solution without fairness is compared to the solution with fairness.

ance/correlation as can be seen in Figure 3.12. The divergence for $\lambda_1 = 4.7$ and $\lambda_2 = 0.0047$ can be solved by choosing a lower learning rate μ. However, one should keep in mind that a lower value of μ will require more iterations to achieve a good convergence for the prediction error. Therefore, a compromise should be found.

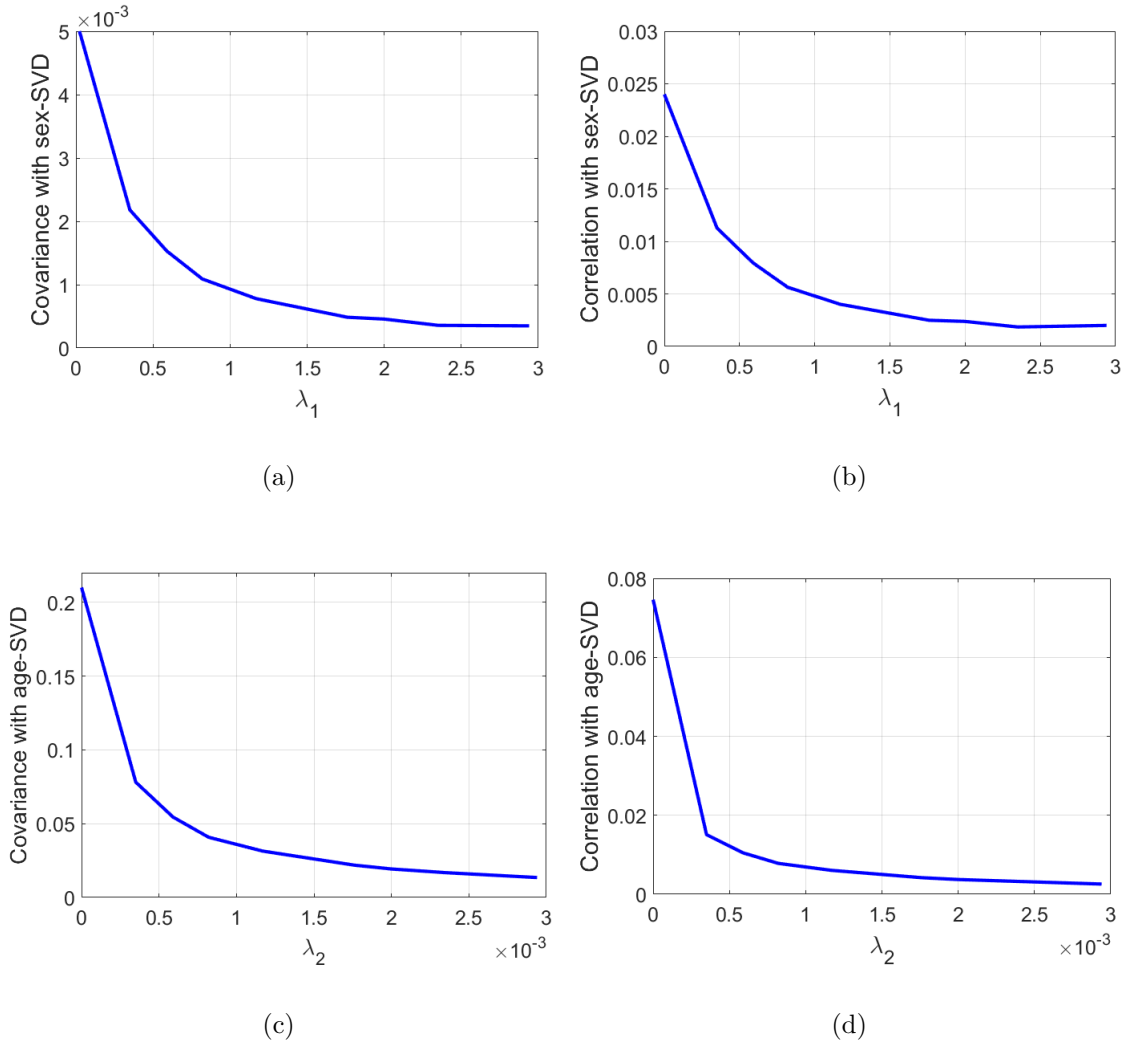
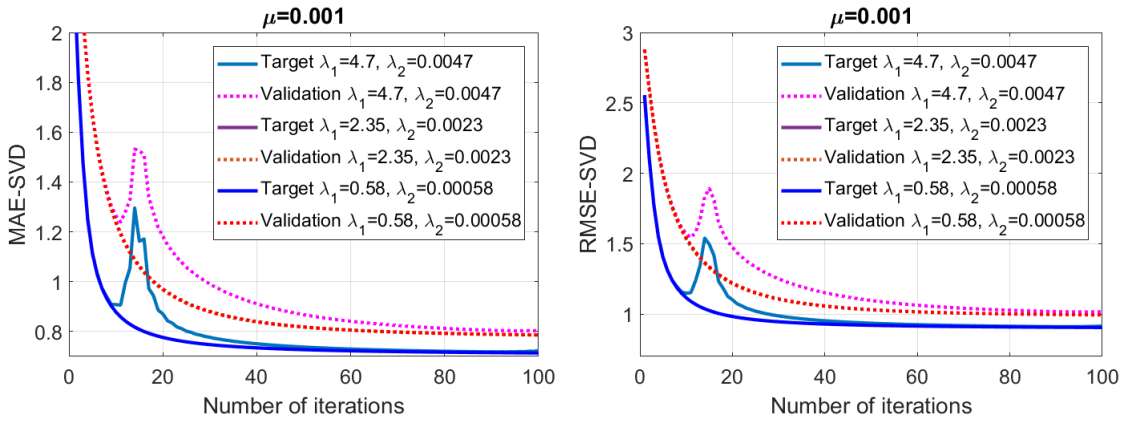


Figure 3.10: Variation of the covariance/correlation as a function of λ.

Finally we have analyzed for each item, the dynamic of the covariance and the correlation before and after prediction. Figure 3.13 shows the results for the sensitive variable "sex". The red curve (covariance/correlation after optimization) is superimposed to the blue one (covariance/correlation before optimization) on the left. Before optimization, the index of the three movies that are the most correlated with the sex are by order, the $220th$, the $699th$, the $278th$, with titles *Mirror Has Two Faces*, *Little Women*, and *Bed of Roses*, respectively. One can also see from Figures 3.13 (a) and (c) that in average, the movies with high index are the ones that are the most correlated with sex. This dynamic is no longer observed after optimization, i.e. in average, the dynamic of the correlation with sex is constant after optimization.

Covariance and correlation with the age, for each item are displayed in Figure 3.14. As for the sex, we can see that movies with high index seem less correlated in the



(a) (b)

Figure 3.11: Illustration of convergence issue on the prediction error, when the learning rate is too high, for a given λ.

initial data set. However, in contrary to the sex case, this dynamic seems preserved after optimization. This difference can be explained by the fact that, in average the covariance terms corresponding to the sex are numerically higher than that of the age. As a consequence, the age part dominates the sex part in the gradient calculation.

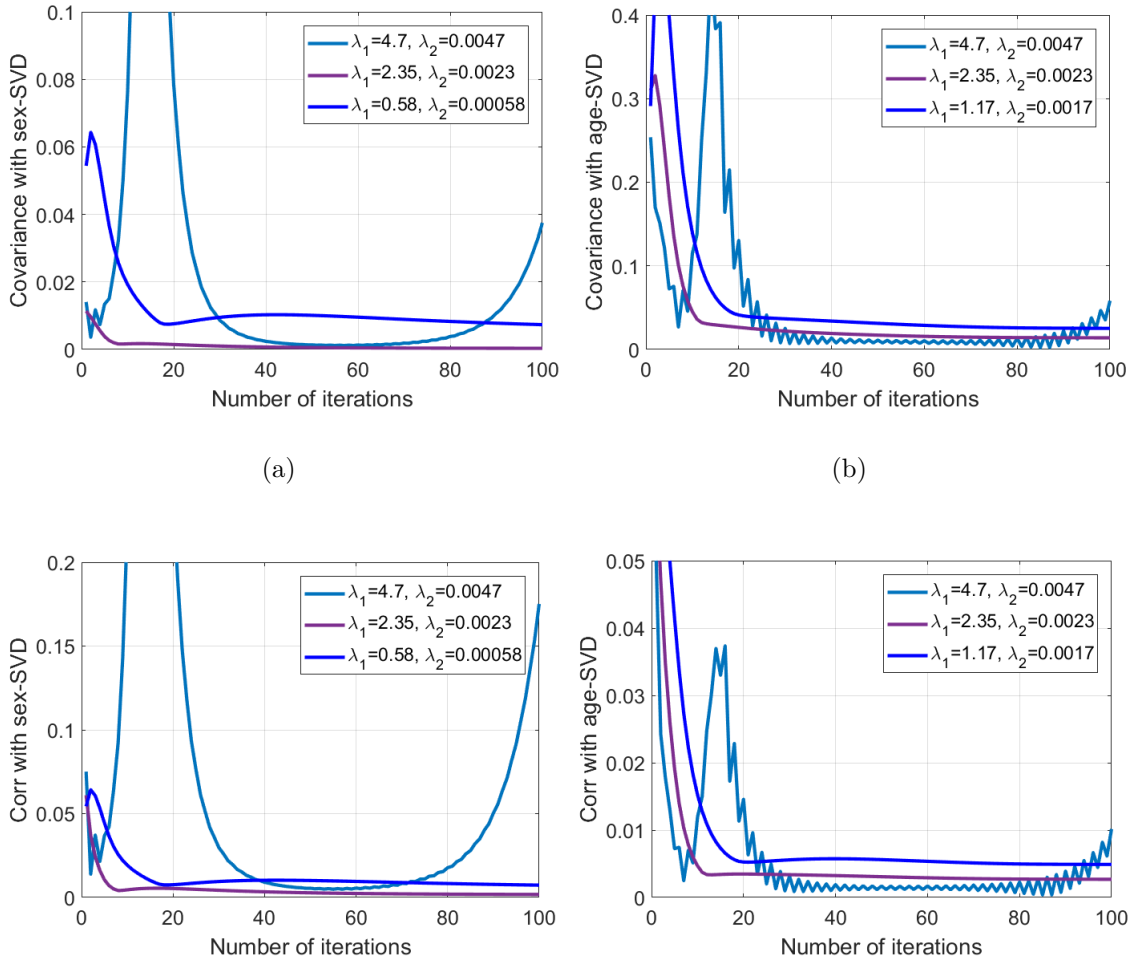


Figure 3.12: Illustration of convergence issue on the covariance/correlation, when the learning rate is too high, for a given λ.

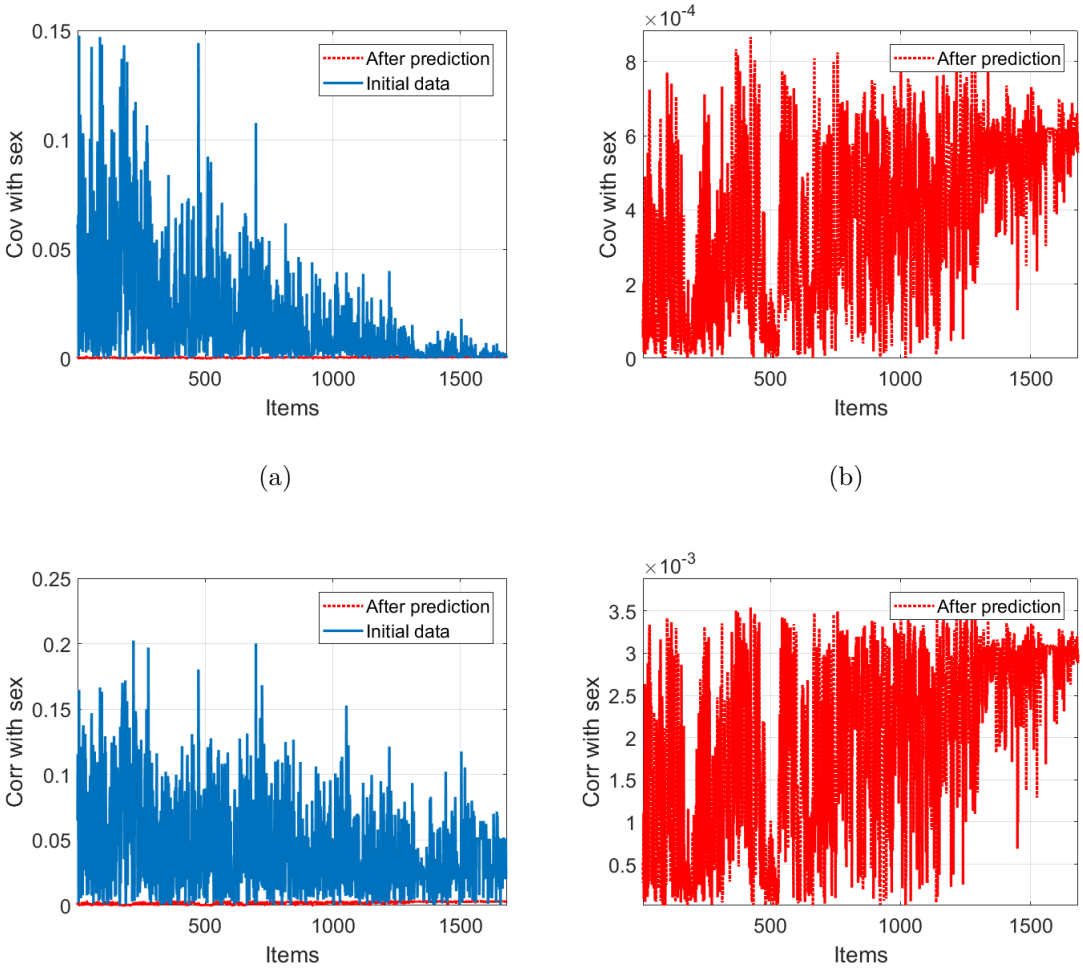
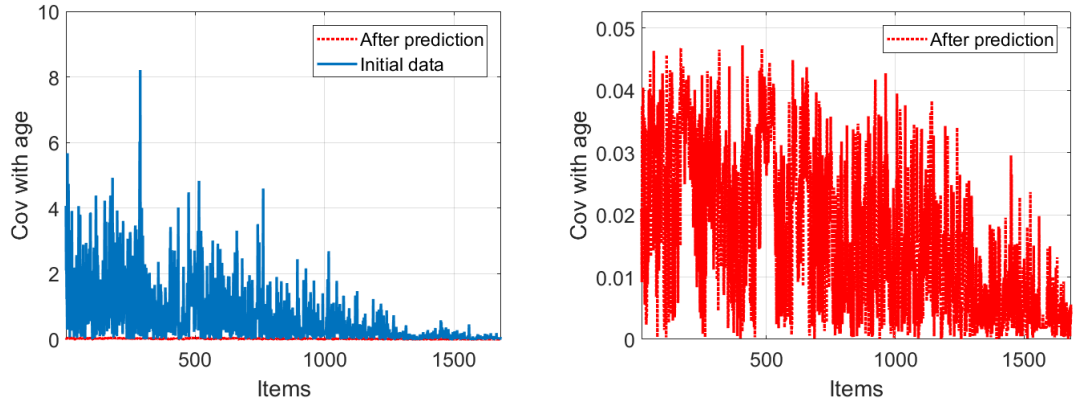
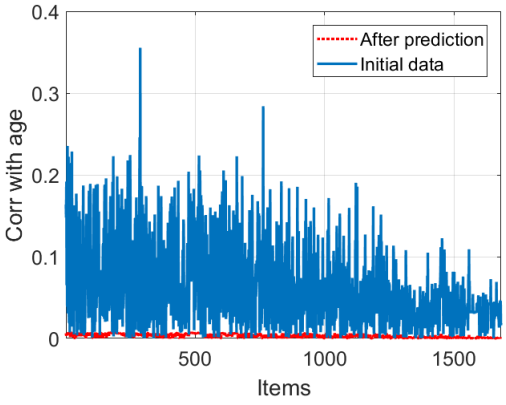


Figure 3.13: Covariance/correlation versus sex, as a function of the item index

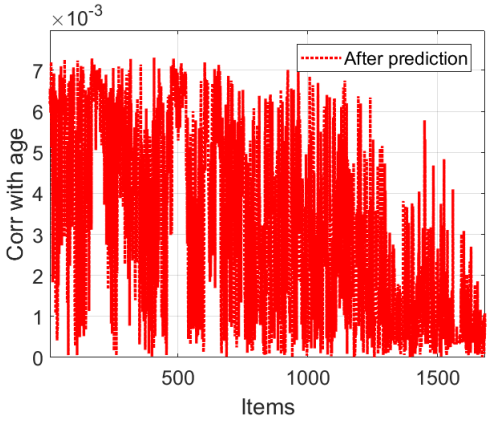


(a)

(b)



(c)



(d)

Figure 3.14: Covariance/correlation versus age, as a function of the item index

4

Conclusion

This thesis has been devoted to the inclusion of fairness in recommendation systems. First, the most intuitive recommendation algorithm, namely the KNN method, has been studied and its performance considering different numbers of neighbors has been illustrated. Then a factorization based method has been analyzed and we have shown that it provides better prediction performance than the KNN method. Thereafter, a parametric study of the factorization based method has been carried out. In particular, we have considered the impact of the factorization matrix size, the impact of using offset variables, the impact of the regularization coefficient, and the impact of the relative size of the training data with respect to the validation data size. We have shown that the obtained prediction error is strongly influenced by the statistics of the chosen validation and training data sets. Furthermore, while the prediction error decreases monotonously as a function of the iteration number, the covariance term modeling the fairness oscillates. As a consequence, depending on the maximum number of iterations at which the algorithm has been stopped, discrimination existing in the initial data set may be (or not) amplified. this observation justifies the need of explicitly including fairness in the algorithm.

In a second part of the work, fairness has been introduced by adding some terms in the objective function to be minimized. Those terms correspond to the covariance between the prediction and a given set of sensitive variables. It has been shown that adding the fairness constraint introduces significant fairness in the prediction while not deteriorating the prediction error. This observation can be explained by the non-convex property of the optimization problem. The impact of varying the weight of the fairness term has been analyzed. We have observed that a bigger weight on the fairness term significantly improves the convergence of the fairness up to a given point. After that, the solution diverges, and one needs to lower the learning rate of the gradient algorithm to obtain a good convergence.

4.1 Future work

First, it has been observed that the performance of the algorithm significantly depends on some parameters: number of iterations, learning rate μ, weight of the covariance term λ, etc. In this work, the values of those parameters have been chosen based on experience, and not from conceptual reasoning. An algorithm able to learn by itself from the data set so as to automatically set those parameters will be very useful. Moreover, here, those parameters have been fixed for all iterations. Being able to adapt the learning rate for example as a function of the local dynamics of the objective function can enable a rapid convergence while not loosing the accuracy of the solution.

Second, the fairness has been modeled as a covariance/correlation with respect to the sensitive variables. This is not the only way to model fairness. Other metrics have been proposed in the literature, either based on the difference of the mean ratings between the group of interests [18], [34], or based on nonparametric statistics tests (Kolmogorov-Smirnov test)[37]. A future work may analyze those metrics, and compare them. For example, it should be interesting to analyze the performance, considering a given metric, when another metric has been used in the objective function, and propose the suitable metrics depending on the application of interest. One can also combine those metrics, for example the correlation metric, with the mean ratings metric, to achieve a good compromise between them.

Finally, mixing collaborative filtering with content based filtering, while preserving fairness can also be a good research topic that will be very interesting in some realistic situations.

References

[1] K. Cukier and V. Mayer-Schoenberger. **The rise of big data: How it's changing the way we think about the world**. *Foreign Affairs*, **92**:28–40, 2013.

[2] S. Barocas and A. D. Selbst. **Big data's disparate impact.** *California Law Review*, **104**:671–732, 2016.

[3] M. Seyedan and F. Mafakheri. **Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities**. *Journal of Big Data*, **7**, 2020.

[4] N. Labi. **Misfortune Teller**. *The atlantic*, 2012.

[5] A. Subrahmanyam. **Big data in finance: Evidence and challenges**. *science Direct*, **19**:283–287, 2019.

[6] M. Yang, X. Shao, G. Xue, and B. Xie. **Big data theory based spectrum sensing algorithm for the satellite cognitive radio network**. *Springer*, 2021.

[7] Y. Huang, J. Tan, and Y. Liang. **Wireless big data: transforming heterogeneous networks to smart networks**. *Journal of Communications and Information Networks volume*, **2**:19–32, 2017.

[8] A. W. Kiwelekar, G. S. Mahamunkar, L. D. Netak, and V. B. Nikamn. **Deep Learning Techniques for Geospatial Data Analysis**. *Springer*, 2020.

[9] D. B. Rawat, R. Doku, and M. Garuba. **Cybersecurity in big data era: from securing big data to data-driven security**. *IEEE Transactions on Services Computing*, 2020.

[10] L. Hardesty. **The history of Amazon's recommendation algorithm**. *amazon.science*, 2019.

[11] D. Chong. **Deep dive into Netflix's recommender system**. *towardsdatascience.com*, 2020.

[12] Z. Cui, X. Xu, F. Xue, X. Cai, y. Cao, W. Zhang, and J. Chen. **Personalized recommendation system based on collaborative filtering for IoT scenarios**. *IEEE Transactions on Services Computing*, **13**:685–695, 2020.

[13] Y. Blanco-Fernandez, J. Pazos-arias, A. Gil-Solla, M. Ramos-Cabrer, and M. Lopez-Nores. **Providing entertainment by content-based filtering and semantic reasoning in intelligent recommender systems**. *IEEE Transactions on Consumer Electronics*, **54**:727–735, 2008.

[14] B. P. S. Murthi and S. Sarkar. **The role of the management sciences in research on personalization.** *Management Science*, **49**:1344–1362, 2003.

[15] R. Conaway and O. Laasch. **Principles of responsible management: global sustainability, responsibility, and ethics.** *Semantic scholar*, 2015.

[16] I. Petrescu. **Social responsibility in modern management.** *Review of General Management*, **28**, 2018.

[17] T. Brennan and W. L. Oliver. **The emergence of machine learning techniques in criminology.** *Criminology Public Policy*, **12**:551–562, 2013.

[18] S. Yao and B. Huang. **Beyond parity: Fairness objectives for collaborative filtering**. *Computing Research Repository*, page arXiv:1705.08804, 2017.

[19] K. D. Johnson, D. Foster, and R. Stine. **Impartial predictive modeling: Ensuring fairness in arbitrary models**. *Statistical Science*, pages 1–29, 2016.

[20] M. Wan, J. Ni, M. Misra, and J. McAuley. **Addressing marketing bias in product recommendations**. *Proceedings of the 13th International Conference on Web Search and Data Mining*, pages 618–626, 2020.

[21] A. Romei and S. Ruggieri. **A multidisciplinary survey on discrimination analysis**. *The Knowledge Engineering Review*, pages 582–638, 2013.

[22] Unites Nations. **Sustainable development goals**. 2015.

[23] O. Harrison. **Machine learning basics with the K-nearest neighbors algorithm**. *Towards data science*, 2018.

[24] Y. Koren, R. Bell, and C. Volinsky. **Matrix factorization techniques for recommender systems**. *Computer*, **42**:30–37, 2009.

[25] C. C. Aggarwal. **Recommender Systems**. *Springer*, 2016.

[26] S. C. Stephen, H. Xie, and S. Rai. **Measures of similarity in memory-based collaborative filtering recommender system: A comparison**. *Proceedings of the 4th Multidisciplinary International Social Networks Conference*, 2017.

[27] C. Ma. **A guide to singular value decomposition for collaborative filtering**. *Computer Science*, 2007.

[28] B. C. Chen, W. Ji, S. Rho, and Y. Gu. **Supervised collaborative filtering based on ridge alternating least squares and iterative projection Pursuit**. *IEEE Access*, **5**, 2017.

[29] D. D. Lee and H. C. Seung. **Learning the parts of objects by non-negative matrix factorization,**. *Nature*, **401**, 1999.

[30] M. Chu, F. Diele, R. Plemmons, and S. Ragni. **Optimality, computation, and interpretation of nonnegative matrix factorizations**. *Dept. Math., North Carolina State Univ., Raleigh, NC, USA, Tech. Rep.*, 2004.

[31] A. Paterek. **Improving regularized Singular Value Decomposition for collaborative filtering**. *Proceedings of KDD Cup and Workshop*, 2007.

[32] J. Bennett and S. Lanning. **The Netflix Prize**. *Proceedings of KDD Cup and Workshop*, 2007.

[33] A. Agarwal, M. Dudik, and Z. S. Wu. **Fair regression: Quantitative definitions and reduction-based algorithms**. *Proceedings of the 36th International Conference on Machine Learning*, 2019.

[34] S. Yao and B. Huang. **New fairness metrics for recommendation that embrace differences**. *Computing Research Repository, arXiv*, 2017.

[35] D. Olivier and K. Brieuc. **Increasing fairness in recommender systems : a study of simple decorrelation applied to the Lecron-Fouss model**. *Master thesis, LSM, UCLouvain*, 2021.

[36] D. Wackerly, W. Mendenhall, and R. L. Scheaffer. **Mathematical statistics with applications**. *Cengage Learning*, 2014.

[37] Z. Zhu, X. Hu, and J. Caverlee. **Fairness-aware tensor-based recommendation**. *Association for Computing Machinery*, 2018.

[38] F. M. Harper and J. A. Konstan. **The MovieLens Datasets: History and Context**. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2015.

Abstract :

Recommendation systems are filtering algorithms developed to predict the preference of a user for a given item. Those systems may be categorized into two groups. In the first group, i.e. collaborative filtering, it is assumed that people who agreed in the past will also agree in the future. The second approach, content based filtering, relies on the expansion of each item into a set of characteristics. In this thesis, the focus is put on the first approach, i.e. collaborative based filtering because this class of methods has the advantage of not requiring an understanding of the items characteristics. However, although collaborating filtering has proven to be able to provide good performance in term of prediction error, those algorithms tend to reproduce or amplify in the predictions possible correlations or discriminations (with respect to some sensitive variables) existing in the training data set. This characteristic may cause in certain circumstances, ethical issues. Therefore, it is in some cases preferable to make the predictions, independent with respect to some sensitive variables so as to introduce more fairness in the recommendation. This additional constraint may (or not) be achieved at the expense of a worse prediction error as compared to the solution with discriminations. The goals of this thesis are twofold. First, two main collaborative filtering algorithms (matrix factorization based method, and the K-nearest neighbors algorithm) are reviewed and compared against each other. Second, fairness constraint is introduced, and its impact on the prediction error is closely analyzed. Ability of the developed algorithm to reduce possible discriminations is demonstrated through various discrimination measures, namely the covariance, and the correlation.

Résumé :

Les systèmes de recommandation sont des algorithmes de filtrage conçus pour prédire la préférence d'un utilisateur pour un objet donné. Ces systèmes peuvent être catégorisés en deux groupes. Le premier groupe, le filtrage collaboratif fait l'hypothèse que les individus qui avaient les mêmes préférences par le passé auront les mêmes préférences à l'avenir. La seconde approche, le filtrage basé sur le contenu, repose sur l'expansion de chaque objet dans une base de caractéristiques. Ce travail se focalise sur la première approche parce qu'elle présente l'avantage de ne pas nécessiter la compréhension des caractéristiques de chaque objet. Cependant, bien que le filtrage collaboratif donne de bonnes performances en terme d'erreur de prédiction, il tend à reproduire ou amplifier des corrélations préexistant dans la base donnée. Cette caractéristique peut dans certaines circonstances poser des problèmes éthiques. Il est donc parfois préférable de rendre les prédictions indépendantes de certaines variables sensibles. Cette contrainte additionnelle pourrait (ou pas) conduire à une dégradation de l'erreur de prédiction. Ce travail a deux objectifs. Dans un premier temps, les deux principaux algorithmes de filtrage collaboratif (méthode basée sur la factorisation matricielle, et l'algorithme des K plus proches voisins) ont été étudiés, et leurs performances ont été comparées. Dans un second temps, la décorrélation avec les variables sensibles est introduite et son impact sur l'erreur de prédiction est analysé. La capacité de l'algorithme à réduire certaines discriminations a été démontrée à travers plusieurs mesures de discrimination : la covariance et la corrélation.

UNIVERSITE CATHOLIQUE DE LOUVAIN
Louvain School of Management
Place des Doyens. 1 bteL2.01.01, 6000 Charleroi, Belgique
Chaussée de Binche 151, 7000 Mons, Belgique
www.uclouvain.be/lsm