Exploiting context-awareness and multi-criteria decision making to improve items recommendation using a tripartite graph-based model

Rim Dridi

IRIT, University of Toulouse 3, Toulouse, France

ENSI, University of Manouba, LISI laboratory, INSAT, University of Carthage, Tunis, Tunisia

Lynda Tamine

IRIT, University of Toulouse 3, Toulouse, France

Yahya Slimani

ISAMM, University of Manouba, LISI laboratory, INSAT, University of Carthage, Tunis, Tunisia

Abstract

Integrating useful input information is essential to provide efficient recommendations to users. In this work, we focus on improving items ratings prediction by merging both multiple contexts and multiple criteria based research directions which were addressed separately in most existent literature. Throughout this article, *Criteria* refer to the items attributes, while *Context* denotes the circumstances in which the user uses an item. Our goal is to capture more fine grained preferences to improve items recommendation quality using users' multiple criteria ratings under specific contextual situations. Therefore, we examine the recommender's data from the graph theory based perspective by representing three types of entities (users, contextual situations and criteria) as well as their relationships as a tripartite graph. Upon the assumption that contextually similar users tend to have similar interests for similar item criteria, we

[⋆]This is a footnote associated with the title.

[∗]This is to indicate the corresponding author.
Email address: rim.dridi@ensi-uma.tn (Rim Dridi)

perform a high-order co-clustering on the tripartite graph for simultaneously partitioning the graph entities representing users in similar contextual situations and their evaluated item criteria. To predict cluster-based multi-criteria ratings, we introduce an improved rating prediction method that considers the dependency between users and their contextual situations, and also takes into account the correlation between criteria in the prediction process. The predicted multi-criteria ratings are finally aggregated into a single representative output corresponding to an overall item rating. To guide our investigation, we create a research hypothesis to provide insights about the tripartite graph partitioning and design clear and justified preliminary experiments including quantitative and qualitative analyzes to validate it. Further thorough experiments on the two available context-aware multi-criteria datasets, TripAdvisor and Educational, demonstrate that our proposal exhibits substantial improvements over alternative recommendations approaches.

Keywords: Recommender systems, multi-criteria decision, tripartite graph, co-clustering, contextual situation, rating prediction.

1. Introduction

Significant research efforts have recently revealed that in addition to designing single-rating systems that produce items recommendations tailored to users' needs, a substantial leap can be taken by integrating richer information in terms of item criteria feedback or users' contexts in the recommendation process [1, 2]. Therefore, traditional single-rating approaches are extended through two research directions to provide a better recommendation performance by making use the available information. The first direction is devoted to context-aware recommender systems (CARS) [2, 3, 4, 5, 6, 7], where useful context-specific dimensions of information (e.g., time, location, etc.) are integrated in the recommendation process. These contextual dimensions affect user's preferences, since user's decisions are always changing from contexts to contexts. One of the well-known approaches in this area is the context-aware matrix factorization

(CAMF) [2], which was initially introduced to embed contextual information in the classic matrix factorization approach for improved recommendations. Several efforts followed the development of CAMF and suggested extended matrix factorization models to adapt them for context-aware recommendations such as [3]. The second direction concerns multi-criteria recommender systems (MCRS) [1, 8, 9, 10, 11, 12], which take advantage of describing an item through multiple criteria and considering the users' feedback on each of them to produce more personalized recommendations. Among the earliest widespread contributions on MCRS, the one proposed in [1] that considers multi-criteria ratings to customize recommendation applications.

This article explores the combination of advantages raisen from the two mentioned research directions. In fact, users tend to give more detailed items feedback based on their criteria ratings and also the contextual situations in which these items are consumed.

As an illustration, consider the popular hotel reviews website (TripAdvisor[1]), through which users can express their stay experience according to hotel criteria (e.g., location, service, cleanliness *etc.*) and their context such as the season of traveling. Obviously, a seaside hotel would be more suitable in summer than in winter leading to consider the location criterion with the temporal contextual dimension to select the convenient hotel.

Despite this interest, nearly all recommendation studies have focused on context-aware or multi-criteria based research directions independently of each other. To the best of our knowledge, only very few studies closest to ours [13, 14] tackled the problem of using context when designing a multi-criteria utility function for measuring the usefulness of items to target users. The contribution in [13] consists in defining a multi-dimensional recommendation space to perform personalized services in mobile commerce based on users neighbors' ratings. This particular recommendation space is represented by a 4-order tensor containing sets of users, items, contexts and criteria. Then, multi-linear singu-

[1] http://www.tripadvisor.co.uk/

lar value decomposition (MSVD) is used to find users' neighbors based on their multi-criteria ratings under relevant context information. In [14], the authors addressed context-awareness and multi-criteria decision making in educational recommender systems. The core idea consists in integrating contextual dimensions (e.g., class, semester) in the multi-criteria baselines steps to enhance the recommendation quality. More specifically, the multi-criteria preferences were viewed as contexts and a context-aware recommendation method was used to integrate both contexts and criteria.

Differently from closely related work [13, 14] in modelling the multi-dimensional available data, we define a new representation modeling the associations between users, contextual situations and criteria as a tripartite graph. To solve the tripartite graph partitioning problem, we create a research hypothesis **H** assuming that users who have common contextual situations tend to have similar criteria interests. According to the posed hypothesis, we explore the idea of triplet data co-clustering embedding contextually similar users providing similar criteria ratings to target items. Precisely, we model the high-order partitioning problem as a consistent fusion of two pair-wise co-clustering sub-problems, with the constraint of the triplet structure. Furthermore, we underline a new challenge through the tripartite graph representation, consisting in weighting the relatedness between the different graph entities.

When it comes to the rating prediction step, we design a new strategy for this task, unlike what was previously performed by closest studies [13, 14] which used traditional existing methods to generate items rating predictions or recommendations. To predict cluster-based multiple criteria ratings for users involved in specific contextual situations, our strategy relies on a novel combination of two recommendation algorithms categories: (1) it first enables the integration of context in the prediction process by considering the dependency between contexts and users in a low dimensional space, and then (2) it emphasizes the inter-dependencies between criteria in the prediction process. Then, the results from the discovered co-clusters including the ratings related to the criteria evaluations can be combined by an aggregation function into a global item rating representing the overall user's assessment of an item.

Finally, we conduct a series of experiments to study the performance of our proposal and to address the following research questions:

(RQ1) What is the correlation strength between contextual situation of users and the importance of criteria they consider for rating an item?

(RQ2) How do our co-clustering context-aware multi-criteria recommendation method perform in comparison to representative baselines ?

The rest of this article is structured as follows: in Section 2, we provide an overview of the existing research on our topic. Section 3 describes the entities of the context-aware multi-criteria tripartite graph as well as their connections. We then introduce the proposed multi-dimensional recommendation model in Section 4. The experimental results are presented and analyzed in Section 5. Finally, we conclude the article and we list some open issues that require future work in Section 6.

2. Related Work

Traditionally, the recommendation is formalized as the problem of predicting the likeliness of an unknown item to a user. Thus, the core algorithm in a recommender system attempts to estimate the utility function $f_R(u, i)$ that measures the appropriateness of recommending an item $i \in Items$ to a user $u \in Users$ as follows:

$$f_R : Users \times Items \to R_0 \tag{1}$$

In the vast majority of recommender systems, the utility function considers two types of entities to produce a single criterion rating (R_0) that represents the overall evaluation of an item by a user in the two-dimensional $Users \times Items$ space. However, this assumption has been considered as limited [1]. Therefore, these systems are extended to provide new lines of research areas such as context-aware recommendation and multi-criteria based recommendation.

2.1. Multi-Criteria Recommender Systems

The recommender systems that focus on describing an item through multiple criteria and consider the users' feedback on them have been referred to as multi-criteria recommender systems (MCRS) [1, 8, 9, 10, 11, 12]. These systems can make more effective recommendations, where users leave not only an overall rating on items, but also multi-criteria ratings that present their preferences on different attributes of items. Accordingly, the utility function $f_R(u, i)$ in MCRS is no longer with a single overall rating (R_0). It additionally considers user's ratings on different item criteria (R_1, R_2, ...,R_k):

$$f_R : Users \times Items \rightarrow R_0 \times R_1 \times R_2 \times ... \times R_k \tag{2}$$

MCRS are often classified as being either heuristic-based or model-based. The heuristic-based methods [1, 9, 10] extend the conventional similarity computation of traditional recommenders to reflect multi-criteria rating information where the users' similarities are computed by aggregating traditional similarities from individual criteria. In particular, the paper in [9] describes an heuristic approach that finds the neighbors of an active user by ranking the item criteria preferences of each user. Then, the obtained ranks are exploited to predict the overall ratings by adapting an extended similarity-based method. In [10], authors proposed an heuristic approach that aggregates the users' similarities and find the overall ratings through genetic programming. On the other hand, in the category of model-based methods [1, 8, 11, 12] to which our proposal belongs, the overall item rating is highly correlated to multi-criteria ratings, where a predictive model is built to estimate the user's overall rating on one item from the observed multi-criteria ratings. The model-based methods are well known for their flexibility and efficiency over heuristic approaches, since they can be applied with any recommendation technique. Besides, the overall item rating is considered simply as just another criterion rating in similarity-based heuristic approaches, while the model-based approach often assumes that the overall item rating serves as an aggregate of multi-criteria ratings. One example is "Criteria Chains", proposed in [8], a model-based approach that constructs a list of criteria as a chain and estimates the user's rating on each criterion one by one using the previous predictions as context information. Afterwards, linear and conditional aggregations are applied for the overall ratings prediction. Another model-based approach is defined in [11], where the authors deal with multi-criteria recommendation by using a unified global and local tensor factorization that learns a global predictive model and multiple local ones. Rashed et al. [12] put forward a non-linear co-embedding model for rating prediction that models the user-item relation as a bipartite graph by leveraging additional criteria and content features using their vector representations.

2.2. Context-Aware Recommender Systems

Context plays a crucial role in recommendation since the information it presents show the status of users in environment, and thus, it affects users' decisions. In this line of research, several researches have been dedicated to context-aware recommender systems (CARS) [2, 3, 4, 5, 6, 7], and such systems are successfully employed to improve traditional recommenders. In CARS, the *contextual dimension* is used to refer to the upper-level contextual concept (e.g., season, trip type). The term *contextual condition* is used to denominate the contextual concept instances, (e.g., winter, summer). Therefore, the *contextual situation* refers to a combination of contextual conditions of different contextual dimensions (e.g., {summer, family trip}). Hence, the initial problem formulation for recommender systems (Equation (1)) is extended as follows:

$$f_R : Users \times Items \times Contexts \rightarrow R_0 \tag{3}$$

We distinguish two categories of work in this area: the first one which is the most widely used, integrates context in a single-criterion based recommendation framework [2, 3, 4, 5, 6, 7] to estimate the relevance of an item according to the current user's context. Adomavicius et al. [5] were one of the first to incorporate the contextual information in recommender systems to make recommendations based on multiple dimensions, profiles, and aggregation hierarchies. In [2], the context-aware matrix factorization (CAMF) approach was initially introduced to contextualize the traditional matrix factorization by incorporating a contextual rating deviation component into the recommendation process. Zheng et al. [3] developed a CAMF approach considering the contextual correlation, which turned to measure the contextual situations similarities by defining three different context similarity models. In [4], another CAMF was proposed to produce points of interest (POIs) recommendations by exploiting user's preferences and real-time demand simultaneously. Cheng and Shen [6], worked on the concept of venue-aware recommender systems for music, and constructed a latent variable model to infer users' music preferences based on the venue and surrounding of users. In [7], Jiang et al. developed an author topic model-based CF method to facilitate comprehensive POIs recommendations for social users. The proposal detects users' similarities based on geo-tags attached to their posts in social networking services.

Recent trends in recommender systems have led to the emergence of a new category that explores the exploitation of both context information and multi-criteria ratings to improve the predictive performance. Only very few studies have focused on combining these two research directions within a single recommender [13, 14]. The main idea in [13] is the expansion of the dimensionality of the recommendation space to produce personalized recommendations based on additional useful information. Particularly, the recommendation space is defined by a 4-order tensor which additionally includes contextual information and multi-criteria ratings in the traditional $User \times Item$ space. Then, user's neighbors are found using the MSVD under relevant contextual information to perform recommendations based on those found neighbors. Recently, authors in [14] integrated context into different previous MCRS and applied them in the area of educational learning. These recommenders have two different methods for each step in the multi-criteria recommendation process. In fact, the independent and dependent methods were used for the multi-criteria rating predictions step, and the linear and conditional aggregation methods for the rating aggregations step. Hence, the context has been incorporated in each method and within each step besides the multi-criteria preferences which were also considered as contexts. Yet, rather than improving the prediction accuracy, the inclusion of context in the process of rating aggregations has caused the performance degradation.

3. The Context-Aware Multi-Criteria Tripartite Graph

3.1. Basic Notations

In what follows, we define the basic terms we are going to use in a clear way. More specifically, these terms are the elements belonging to the matrix representation M of the multi-dimensional $Users \times Items \times Contextual\ Situations \times Criteria$ recommendation space.

- Items.

  An item is the general term used to denote what the recommender system suggests to users. An item could be a movie to watch, text to read, product to buy or anything else depending on industries. We denote by I the set of items $I = \{i_1, .., i_z\}$, where z is the total number of items.

- Users.

  A user u may rate one or more criteria of an item in a specific contextual situation. Let U be the set of users $U = \{u_1, .., u_p\}$, where p is the total number of users.

- Contextual situations.

  In recommendation systems area, the contextual situation is defined as a set of contextual dimensions values that describe the context in which the user consumed the item. Formally, let contextual dimensions set, noted Cd is represented by $Cd = \{cd_1, .., cd_k\}$, where k is the total number of contextual dimensions.

  Example. In tourism domain, Cd could include the following contextual dimensions: Cd={Trip type, season}. The contextual conditions set Cc_i of a specified contextual dimension Cd_i is represented by $Cc_i = \{cc_{i1}, .., cc_{il}\}$, where l is the total number of the contextual conditions corresponding to

Cd_i with $1 \leq i \leq k$.

Example. The first contextual dimension (Cd_1: Trip type) is defined by the following contextual conditions Cc_1={Family trip, solo trip, couple trip, friends trip}. A contextual situation is built up as an entity noted s_j, defined as a vector formed by the contextual conditions of k contextual dimensions $s_j = \{cc_{1j}, .., cc_{kj}\}$, where $1 \leq j \leq m$ leading the whole set of situations noted as $S = \{s_1, .., s_m\}$.

Example. A contextual situation s_1 can be defined by the first contextual conditions of both trip type (cc_{11}: family trip) and season contextual dimensions. s_1={ family trip, summer}.

- Criteria.

The criteria set consists of rated item aspects in different contextual situations. It is noted $C = \{c_1, .., c_n\}$, where n is the number of criteria taken in regard for rating an item.

Example. For the item "hotel", the criteria set can be defined by $C =$ {location, service, quality of room}.

3.2. The Graph Structure

To deal with the multi-dimensional available data, we present the context-aware multi-criteria network by a weighted tripartite graph G_{SUC}=(S, U, C, E^{SU}, E^{UC}) where S, U, C stand for the finite sets of contextual situations, users and criteria vertices; E^{SU} and E^{UC} denote the two types of edges in the network that express relationships between user-contextual situations and user-criteria respectively. The tripartite graph G_{SUC} is projected into two bipartite graphs denoted as the contextual situation-user bipartite graph and the user-criteria bipartite graph represented respectively by: G_{SU}=(S, U, E^{SU}) and G_{UC}=(U, C, E^{UC}). More precisely, in the first graph G_{SU}, the edge (s_i, u_j)$\in E^{SU}$ is the undirected link between the contextual situation $s_i \in S$ and the user $u_j \in U$ where $w_{ij}^{(su)}$ denotes the corresponding weight edge. In the second graph G_{UC}, a user u_j and a criterion c_o are connected by an association if u_j has rated c_o. This relationship is established in one direction and is modeled by an edge (u_j,

$c_o) \in E^{UC}$ where the weight of such association is denoted by $w_{jo}^{(uc)}$. Graph entities relationships are represented by weighted edges that model the links between entities.

The two defined types of edges are weighted as follows:

Contextual situation-user associations weighting: To assign the importance weight of each edge $(s_i, u_j) \in E^{SU}$, we use the popular weighting scheme TF-IDF (Term Frequency Inverse Document Frequency) [15], which to date has tended to focus on assigning weight edges of bipartite graphs rather than tripartite graphs. This choice is motivated by TF-IDF's ability to reveal latent connections between users and their contextual information while rating the items.

However, a contextual sparsity problem can be emerged when the total number of contextual conditions would increase exponentially over the contextual dimensions. Therefore, we represent user's contextual information by a contextual situation including a set of contextual conditions. This is a fundamental step to further discover similar contextual groups. Besides, the graph based modeling could assist to reduce the sparsity issue due to its nice structural properties in discovering the hidden relationships and simulating the preferences propagation. To apply the TF-IDF, a document is considered as a user and a term as its contextual situation. As a result, we obtain a SF-IUF (Situation Frequency Inverse User Frequency) that is defined as follows:

$$w_{ij}^{(su)} = SF(i,j) \times IUF(i). \tag{4}$$

$$SF(i,j) = \frac{freq(i,j)}{\max[freq(j)]} \tag{5}$$

Where, $freq(i,j)$ is the frequency of a contextual situation s_i in which the user u_j rated an item and $max[freq(j)]$ is the maximum frequency computed over the frequencies of all situations where user u_j is involved in.

$$IUF(i) = \frac{|U|}{|U_i|} \tag{6}$$

To find the Inverse User Frequency IUF(i), we need to divide the total number of users $|U|$ by $|U_i|$ which is the number of users in the contextual situation s_i. Then, we define a weighted adjacent matrix A for the bipartite graph G_{SU}, element a_{ij} of matrix A can be defined as:

$$a_{ij} = \begin{cases} w_{ij}^{(su)} & \text{if } (s_i, u_j) \in E^{SU} \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

User-criteria associations weighting: The weight of such association $w_{jo}^{(uc)}$ between a user u_j and a criterion c_o is computed as an average rating $\bar{r}_{jo}$ of u_j for c_o across all items I:

$$w_{jo}^{(uc)} = \bar{r}_{jo} = \sum_{t \in I} \left(\frac{r(j,t,o)}{I}\right) \tag{8}$$

Where $r(j,t,o)$ is the rating given by u_j for c_o of the item t.
The weighted adjacent matrix $B = \{b_{jo}\}$ for the bipartite graph G_{UC} can be written as follows:

$$b_{jo} = \begin{cases} w_{jo}^{(uc)} & \text{if } (u_j, c_o) \in E^{UC} \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

Example *(Hotel recommendation)*

Suppose there are three travellers: John, Jack and Amy who tend to select an hotel according to their contextual situations and the hotel criteria: location, service and quality of room. John is going on a business trip that will take place from 9-11 December. Amy and Jack are a couple going on summer vacation. In this example, we can define the users set $U = \{John, Jack, Amy\}$, the contextual situations set $S = \{\{business\ trip, winter\}, \{family\ trip, summer\}\}$ and the criteria set $C = \{location, service, room\}$. The Figure 1 shows the tripartite graph constructed based on this example.

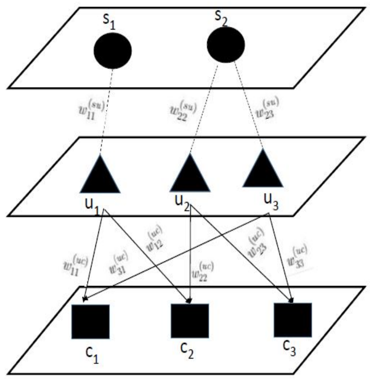


Figure 1: Example of tripartite graph structure

4. The Context-Aware Multi-Criteria Recommendation Model

The pending problem we address consists in predicting users' item ratings in accordance with their contextual situations and criteria preferences. In particular, we turn this problem into two main sub-problems:

(1) identifying co-clusters in the context-aware multi-criteria tripartite graph G_{SUC} by exploiting the edges weights (Section 4.1);

(2) predicting users' preferences on items from estimating the items multi-criteria ratings (Section 4.2).

The presented sub-problems are very challenging. In fact, in the first one, the tripartite graph includes three different kinds of entities, which makes it difficult for traditional clustering methods to accurately find meaningful subgroups simultaneously. For example, it was proved in [16] that extending the spectral graph partitioning method to the high-order case can not really provide the desirable co-clustering results. The second sub-problem faces the rating prediction problem on each criterion in a multi-dimensional recommendation space, where the criteria dependency is often ignored in the prediction process.

4.1. Tripartite Graph Co-clustering

To solve the tripartite graph partitioning problem, we rely on the following driving hypothesis **H** that guides our investigation:

"H: Users in similar contextual situations tend to have similar interests for similar criteria".

As announced in **H**, the involved triplet data consists of contextual situations S, users U and criteria C which represent the multi-type entities in our tripartite graph G_{SUC} described in Section 3.2. To perform the triplet data co-clustering, G_{SUC} will be treated as two dependent bipartite graphs denoted as G_{SU} and G_{UC} which share the central type U. Therefore, the high-order co-clustering problem is modeled as two pair-wise partitions for the sub-problems of S-U co-clustering and U-C co-clustering with the constraint of the triplet structure. The aim of this partitioning is to obtain the desired co-clusters of contextually similar users evaluating similar criteria. In this respect, we employ the Collective Matrix Factorization (CMF) algorithm [17] which is proved to be successful in mining multi-relational data over other co-clustering algorithms such as spectral relational clustering. Broadly speaking, the CMF addresses the problem of simultaneously factoring related matrices describing the connections between their entities that are represented by graphs. The use of CMF can be viewed as a performing co-clustering method to identify both row and column entities clusters simultaneously. In our case, we consider the two related adjacency matrices defined as $A \in R^{m \times p}$ and $B \in R^{p \times n}$ in Section 3.2, which characterize the two dependent bipartite graphs G_{SU} and G_{UC} respectively, where m is the number of contextual situations, p is the number of users and n is the number of criteria. These data matrices involve the three considered graph entities (S, U, C) and their association strengths, sharing the same entity U. For CMF, low-rank matrix factorization is considered as its building block, which extends factoring one matrix to factoring sets of related matrices, i.e., matrices which shared entities. As a result, the co-clustering process using CMF consists in decomposing each adjacency matrix into the product of two matrices to simultaneously obtain subgroups of its rows and columns. Hence, we jointly decompose the input matrices A and B into three lower dimensional matrices that we call T, W, and Z for inducing a co-clustering of entities. More precisely, we approximate each matrix by product of low rank-K factors that form the representations of the associated row and column entities.

The $m \times p$ situations-users matrix A is represented as a function of the product of two lower-rank factor matrices: an $m \times k$ matrix T, and a $p \times k$ matrix W. That is, $A \approx f(TW^T)$ for an element-wise transformation $f\colon R \rightarrow R$ and $k < \{m, p\}$.

Since the matrix B shares the same entity type U with the matrix A, we use the factor W in both approximations and thus we have:

The $p \times n$ users-criteria matrix B is represented as a function of the product of two lower-rank factor matrices: an $p \times k$ matrix W, and a $n \times k$ matrix Z. That is, $B \approx f(WZ^T)$ for an element-wise transformation $f\colon R \rightarrow R$ and $k < \{p, n\}$.

So, each entity-type has a low-rank representation where $T \in R^{m \times k}$ represent the obtained contextual situations factors, $W \in R^{p \times k}$ represent users factors, $Z \in R^{n \times k}$ represent criteria factors and f is a link function.

These obtained factors must be used as arguments of the losses measuring how close the input matrices (A and B) are to their reconstructions: $A \approx f(TW^T)$ and $B \approx f(WZ^T)$. To share information between the two related input matrices, we tie their losses to express the objective function that is being sought to be optimized based on constraints including non-negativity and stochasticity of rows, i.e., the factors are non-negative and normalized. In addition, we include a regularization term to help optimizing the collective factorization model.

Accordingly, the co-clustering problem is modeled as an optimization process where the objective function to minimize is the sum of the reconstruction losses for A and B plus a regularization term $R(.)$:

$$\underset{T,W \geq 0}{\arg\min}\, f = \alpha D(A, TW^T) + (1-\alpha) D(B, WZ^T) + R(T, W, Z) \qquad (10)$$

More clearly, $D(A, TW^T), D(B, WZ^T)$ are the loss functions quantifying the goodness of the approximations and $\alpha \in [0, 1]$ represents a trade-off parameter to weight the relative importance between two relations. Collective matrix factorization assumes that the loss is decomposable. For instance, for the model $A \approx f(TW^T)$ the loss $D(A, TW^T)$ decomposes into a weighted sum over the elements of A. For example, the loss for weighted singular value decomposition is: $D(A, TW^T) = ||W_e \odot (A - TW^T)||^2$. Where W_e is an argument of the loss representing the data weights and $\odot$ denotes the element-wise product of matrices.

We have added the regularization penalty $R(T, W, Z)$ to the objective in order to mitigate overfitting. Standard regularizers for linear models, such as the l_p norms of the factors, can be adapted. In our experiments, we used l_1 norm based regularization which has the merits of being robust and decomposable: $R = \gamma_1||T||_1 + \gamma_2||W||_1 + \gamma_3||Z||_1$, with $\gamma_1, \gamma_2, \gamma_3$ are the regularization control parameters.

After defining the objective function, we need to solve the optimization problem. This resultant problem requires differentiating the objective with respect to each of the obtained factors T, W and Z. Since the loss is a linear function of individual losses, we derive an efficient Newton update using stochastic constraints for finding the roots of the differentiable functions. Here, the optimization process is used to determine optimal co-clusters of the graphs entity-types, cyclically till convergence. Thus, the optimisation of (10) leads to find optimal partition of data driving to a simultaneously clustering of the connected graphs entities into T co-clusters $C_l = \{c_{l_1}, .., c_{l_T}\}$.

4.2. Rating Prediction

Here, we introduce an improved rating prediction method which runs in two stages: (1) predict users' criteria ratings within similar contextual situations; (2) compute users' overall ratings on items.

4.2.1. Criteria Rating Predictions

In this step, we introduce an improved rating prediction method to provide, as an output, the criteria predicted ratings in each co-cluster $c_{l_k}, k \in \{1, .., T\}$ of contextually similar users with preferred criteria. Accordingly, we exploit the co-clusters set C_l derived from the tripartite graph G_{SUC} partitioning process (Section 4.1) which takes into account A and B matrices representing the considered triplet graph entities (S, U, C). Precisely, the object of this step is twofold:

(a) First, the contextual dimensions values representing the contextual situations are fused to users providing a reduced recommendation space by taking into account the dependency between users and their contextual situations. Thus, the contexts are eliminated from the original matrix M representing the multi-dimensional $U \times I \times S \times C$ recommendation space which will be transformed to a new matrix denoted as R representing a lower dimensional $U \times I \times C$ recommendation space. We adopt the user splitting approach [18] for dealing with such problem, where a user may be considered as multiple users, if he or she demonstrates significantly different preferences in different contexts. To better understand the user splitting process, we consider the following educational recommendation example:

User	***Item***	***Rating***	***Application***	***Data***	***Ease***	***Class***	***Year***	***Semester***
st_1	tp_1	4	4	4	4	DM	2017	Fall
st_1	tp_2	2	2	2	2	DA	2017	Spring
st_1	tp_3	4	4	5	4	DA	2018	Fall

Table 1: Example of the Educational context-aware multi-criteria rating dataset

Table 1 presents an example of the Educational dataset [19] that will be used later in our experimental study. In this example, there are one student (st_1), three topics of projects (tp_1, tp_2 and tp_3), three criteria (i.e., application, data and ease) and three contextual dimensions: the type of the class (i.e., database (DB), data analytics (DA) and data mining (DM)), semester (Spring or Fall) and year of the course (2017 or 2018). The user splitting tries to find a contextual condition on which to split each user. In our case, the contextual condition selection process is done by measuring the significance of rating differences given by the student st_1. Impurity criteria [20] are used here to determine how much

st_1 provided different ratings in the presented contextual conditions. Assume that the best split for student st_1 is "semester = fall vs spring", st_1 can be split into two new ones : st_{11} (st_1 choose a project topic for the fall semester) and st_{12} (st_1 choose a project topic for the spring semester). As a result, in this example, the contexts are excluded from the original matrix presented in Table 1 which is converted to a lower dimensional matrix shown in Table 2.

User	***Item***	***Rating***	***Application***	***Data***	***Ease***
st_{11}	tp_1	4	4	4	4
st_{12}	tp_2	2	2	2	2
st_{11}	tp_3	4	4	5	4

Table 2: Example of the transformed Educational context-aware multi-criteria rating dataset

(b) Second, we perform for the second goal a rating prediction algorithm that sketches the dependencies between correlated criteria. Previous recommendation studies generally ignore the correlation aspect when producing rating predictions, not to mention the contributions in the area of multi-criteria recommendation. As far as we know, only one recent research [14] has considered the dependency among multiple criteria to make predictions for users' ratings on these criteria. However, in some cases the used dependent-based criteria method showed worse performance results than those achieved without considering criteria dependency.

We present in Algorithm 1 our criteria rating prediction process that takes as input parameters the converted rating matrix R, the obtained co-clusters set C_l and the factors number F. The presented algorithm aims to provide, as output, the users criteria predicted ratings in each co-cluster $c_{l_k} \in C_l$ by considering criteria dependency. As stated in Algorithm 1, we begin by partitioning the transformed rating matrix R into more specialized sub-matrices according to the clustering results. More precisely, for each co-cluster k (c_{l_k}) containing contextually similar users evaluating preferred criteria, we can get a sub-matrix R_k from the matrix R formed by only rows and columns of users and criteria ap-

Algorithm 1: Criteria Rating Prediction in each Co-cluster

Input: Converted multi-criteria rating matrix R, all the co-clusters $\{c_{l_1}, .., c_{o_T}\}$, and the number of factors F.

begin

  for *each co-cluster* $k \in \{1, .., T\}$ **do**

    R_k=ExtractSub-matrix (R, c_{l_k})

    P_k,Q_k=MatrixFactorization(R_k,F)

    for *each* $j \in P_k$ **do**

      for *each* $t \in Q_k$ **do**

        for *each* $f \in \{1, .., F\}$ **do**

          $Corr(c_k, c_E)$=1-$EuclideanDist(c_k, c_E)$

          $\hat{r}_{j,t,c_k} = p_{j,f} \times q_{t,f} \times Corr(c_k, c_E)$

  Output: Criteria predicted ratings in each co-cluster

pearing in that co-cluster. Then, we apply the Matrix Factorization (MF) [21] for decomposing each obtained sub-matrix R_k into the product of two lower dimensional matrices. The first one is called P, where each row of P would represent the strength of the associations between a user and the features. The second matrix is called Q, where each row of Q would represent the strength of the associations between an item and the features. P and Q are learned using stochastic gradient descent method (SGD) [22] by minimizing the rating prediction errors. After decomposing the sub-matrices, we apply an efficient correlation-based rating prediction algorithm [3]. In this approach, the considered correlation concerns the contexts and it is called "contextual correlation", where the similarity between contextual situations were measured. The core idea behind the notion of "contextual correlation" is that recommendation lists should be similar if their contextual situations are correlated or similar. According to our contribution, the function $Corr(c_k, c_E)$ estimates the correlation between the current criterion c_k (for a user j and an item t) and an unknown criterion c_E. In our model, we assume that the criteria form a multidimensional coordinate system, so that each criterion can situate a position in the corresponding axis. Accordingly, the distance between two such points can be used as the basis for a correlation measure. In our experiments, we use the Euclidean distance to measure the distances. To make sure the resulting correlation values are in the range [0, 1], the values assigned to the criteria were normalized. Finally, in line 4, the correlation-based prediction algorithm [3] can be formulated as follows:

$$\hat{r}_{j,t,c_k} = \vec{p_j}.\vec{q_t}.Corr(c_k, c_E) \tag{11}$$

Where $\vec{p_j}$ represents a weighted user vector, and $\vec{q_t}$ a weighted item vector. The user and item vectors, as well the positions of each criterion are the parameters to be learned by SGD method to minimize the rating prediction errors.

4.2.2. Item Rating Prediction

After predicting the multi-criteria ratings, we need to aggregate them together into a single representative output corresponding to an overall item rating. Hence, this latter is not just another independent rating, but rather serves as an aggregation of the multi-criteria ratings of a given item. In fact, it is not sufficient to only predict criteria ratings, as one of the main goals of recommendation systems is to be able to predict the overall rating of each item for each user, which is beneficial in several situations. Indeed, MCRS ultimately require to compare the items in terms of their overall ratings and recommend only the most pertinent ones. In contrast, to find the best items for users in the absence of the items overall ratings, the recommender will face an extremely complicated multi-criteria optimization problem. Thus, defining the aggregation function is significant for multi-criteria recommendation. As shown in Equation 12, the aggregation function f_R represents the relationship between the multi-criteria ratings $(R_1, R_2, ..., R_k)$ and the overall rating R_0 in order to aggregate these criteria ratings for predicting how much a user will prefer an item.

$$R_0 = f_R(R_1, R_2, ..., R_k) \tag{12}$$

For this task, various aggregation functions have been used. Some of these functions are not appropriate (e.g. average function) as they do not reveal the optimal weights of the item criteria. To generate the function f_R, we use the linear aggregation [1] which can be written as:

$$R_0 = w_1 * R_1 + w_2 * R_2 + ... + w_k * R_k + c \tag{13}$$

Mathematically, this technique assumes that there is a linear relation between the various criteria and the user's overall assessment of an item. More specifically, we suggest the use of multiple linear regression for determining the R_0, where w_o is the weight associated with criterion c_o representing the importance of this criterion. Technically, regression-based techniques have been proven to be promising when predicting the user's overall rating of an item from its criteria ratings [23]. To achieve better performance, we choose the linear-aggregation based multi-criteria recommendation method "Support Vector Regression (SVR)" [24]. One reason for this choice is that SVR has been proven to work well in these aggregation-based approaches by showing higher predictive accuracy and handling very sparse datasets. This is typical, for example, in the tourism domain.

The parameters in the linear aggregation (shown in Equation 13) such as the criteria weights $(w_1, w_2, ..., w_k)$ as well as the constant c are learned by minimizing the squared prediction errors during the training.

5. Experimental Evaluation

In this section, we introduce the two main parts of the evaluation experiments including the experimental setting and the comparison results and discussion.

5.1. Experimental Evaluation Setting

5.1.1. Datasets

One well-known difficulty of recommendation research is to find datasets with both multi-criteria preferences and users' contexts. To the best of our knowledge, there are only the two following datasets, that we used, providing multi-criteria users' preferences in context:

- **TripAdvisor dataset** [24]: includes 3,524 ratings given by 551 users on 3,716 hotels. Note that we use as contextual dimensions the *trip type*, the *year* and the *season* which is derived from the trip month. The dataset contains ratings of 7 individual item criteria (value, location, rooms, cleanliness, check-in, service and sleep quality), plus 1 overall rating. The TripAdvisor dataset also includes other information about users (e.g., member type) and items (e.g., item type, amenities, item locality). A 5-star rating scale is used in this dataset, ranging from 1 ("terrible") to 5 ("excellent"). On average, every user has provided about 3 rated hotels in the dataset, which satisfies the experiment requirement.

- **Educational dataset** [19]: includes 3,306 ratings given by 269 students on 70 items (i.e., the topics of the projects). In the Educational dataset, users are all students but they have different profiles. In fact, the user information in this dataset includes demographic information about each student, such as age, gender, nationality, marriage status, and the personality traits. The dataset also contains items overall ratings and ratings on 3 different criteria: (i.e., application, data and ease). Each student was asked to select at least 6 topics of the projects, and provide an overall rating to them. There are 3 contextual dimensions: the *type of the class*, the *semester* and the *year of the course*. All the criteria ratings as well as the overall ratings ranged from 1 to 5, where 1 indicates the less favorite and 5 the most favorite.

5.1.2. Evaluation Protocol

On the above datasets, we used a five-fold cross-validation technique for both parameter tuning and evaluation. We used the Mean Absolute Error (MAE) metric to study the effects of the used hyperparameters on the ratings prediction accuracy. To measure the top-N recommendation task, we relied on the rank-based Precision and Recall metrics where each metric is computed per user and then averaged across all users. Precision is the fraction of positive predictions over the top-N recommendations in a specific context, while Recall refers to the fraction of positive predictions with respect to the items items recommended in a specific context and situated in the top-N positions. However, the Precision results can increase at the cost of the Recall results sometimes. So, as commonly adopted in the literature ([14, 25]), we consider the popular harmonic F-score measure that combines the Precision and Recall metrics together to offer a more balanced view to a recommender system performance. Therefore, F-score can be represented by Equation 14.

$$F - score = \frac{2.Precision.Recall}{Precision + Recall} \tag{14}$$

In the comparison results, we present the F-score results based on the top-10 recommendations, since the results on the top-5 recommendations display the similar patterns. For this purpose, we used CARSKit [26] contextual recommendation library in order to produce intermediate results in the recommendation process.

5.1.3. Baselines

We use baselines that fall in four categories of conventional state-of-the-art recommendations algorithms. The first one includes single rating based methods which only uses a single criterion for item rating (`BiasedMF`). The second category contains a context-aware based rating method which integrates context information into a traditional recommender process (`CAMF`). The third category concerns the multi-criteria rating based methods which consider multiple criteria and take into account the users' feedback on them (`ABM`, `CIC`, `CCA`, `CCC`). The last category considers the context-aware multi-criteria rating based methods (`DCL`, `CIL`, `CCIC`, `CDL`, `DCC`). We now briefly describe these baselines:

- `Biased Matrix Factorization (BiasedMF)` [21]: a matrix factorization model without context and multi-criteria ratings.

- `Context Aware Matrix Factorization (CAMF)` [2]: a matrix factorization model that integrates the context information only.

- `Aggregation-based Model (ABM)` [1]: uses independent criteria rating predictions and the linear aggregations.

- `Criteria-Independent Contextual Model (CIC)` [8]: uses independent criteria rating predictions and the conditional aggregations.

- `Criteria Chains:Aggregation Model (CCA)` [8]: uses dependent criteria rating predictions and the linear aggregations.

- `Criteria Chain:Contextual Model (CCC)` [8]: uses dependent criteria rating predictions and the conditional aggregations.

- `Criteria-Dependent Contextual Linear Model (DCL)` [14]: uses dependent criteria rating predictions and contextual linear aggregations.

- `Contextual Criteria-Independent Linear Model (CIL)` [14]: uses independent contextual criteria rating predictions and linear aggregations.

- `Contextual Criteria-Independent Conditional Model (CCIC)` [14]: uses independent contextual criteria rating predictions and conditional aggregations.

- `Criteria-Contextual Dependent Linear Model (CDL)` [14]: uses contextual dependent criteria rating predictions and linear aggregations.

- `Criteria-Dependent Contextual Conditional Model (DCC)` [14]: uses dependent contextual criteria rating predictions and conditional aggregations.

- `Criteria-Independent Contextual Linear Model (ICL)` [14]: uses independent criteria rating predictions and contextual linear aggregations.

- `Criteria-Independent Contextual Conditional Model (ICC)` [14]: uses independent criteria rating predictions and contextual conditional aggregations.

It worth of mention that the baselines experimental results presented in the comparison study (section 5.2.3) are obtained as follows:

- For the two first categories including single and context-aware rating approaches, we used the corresponding available models (BiasedMF and CAMF) implemented in the recommendation library CARSKit [26].

-For the two latter categories including multi-criteria and context-aware multi-criteria rating approaches, their corresponding models are not provided in the recommendation library CARSKit. Thus, we reported the results published in papers [8] and [14] with respect to the measures and datasets used in those papers. Accordingly, improvements are computed when comparable results using the same metrics and datasets are available.

5.1.4. Configurations

To perform the criteria ratings prediction step, we first need to examine which contextual dimensions can be used as the best split. For this purpose, we used the impurity criterion t_{mean} [20] which estimates the statistical significance of the difference in the means of ratings of each alternative contextual condition using a t-test. Usually, a threshold for the splitting process should be fixed so that users are only be split when the splitting criteria meets the significance requirement. So, we use a threshold that approximately corresponds to the 0.05 level of statistical significance. Once the splitting has been performed, we use the correlation-based CAMF defined in Equation (11) on the resulting matrices and use the factors for computing item criteria ratings in each co-cluster. In order to better evaluate the correlation-based CAMF, we tried a range of different latent factors F ($5 \leq F \leq 60$, increment 5) and training iteration It ($10 \leq It \leq 100$, increment 10). Other parameters like learning and regularization factors are handled by CARSKit [26], where SGD is used as the optimization method.

5.2. Results and Discussion

5.2.1. Analysis of the relationship between contextual situations and criteria importance (RQ1)

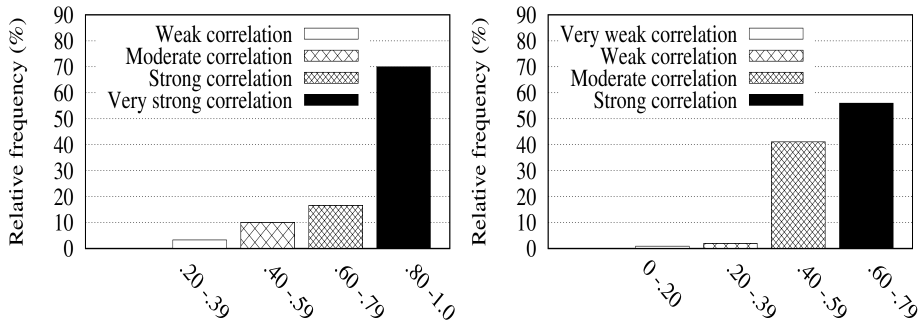
Quantitative analysis. The aim of this quantitative study is to determine the strength and the significance of the correlations between contextually similar users with respect to their criteria importance. We begin by computing users' criteria importance to identify users' preferred criteria depending on their contextual situations. More specifically, we incorporate user's contextual situation in finding users' preferred criteria using the formula described in [9] as follows:

$$Pref_{u,s}^{c} = \frac{\sum_{i \in I_{u,s}} n_{i,s} \times \Gamma_{u,i,s}^{c}}{\sum_{i \in I_{u,s}} n_{i,s}} \tag{15}$$

$$\Gamma_{u,i,s}^{c} = r_{u,i,s}^{c} - \bar{r}_{i,s}^{c} \tag{16}$$

Where $I_{u,s}$ is the items set rated by u in the contextual situation s and $n_{i,s}$ is the number of users who evaluated item i in the same contextual situation s. $\Gamma_{u,i,s}^{c}$ denotes the tendency of rating a criterion c of a rated item i by user u in s, where, $r_{u,i,s}^{c}$ is the criterion rating provided by u of the item i in s and $\bar{r}_{i,s}^{c}$ is the average rating.

Thereafter, we study the strength of users' relationships with respect to their criteria importance by the frequently used Spearman's rank correlation coefficient (r_s) [27]. In Figure.2 the users' relative frequency in each correlation strength interval and the significance testing based on the Student t-test statistic are computed for both TripAdvisor and Educational datasets. As shown in Figure.2a, for the TripAdvisor dataset, the highest percentage (70%) corresponds to the very strongly related users in similar contextual situations, where the Spearman's correlation coefficient falls into the interval [0.80, 1]. We also notice in Figure.2b an important percentage of strongly related contextually similar users that reaches 56% in the Educational dataset. These results are also found to be statistically significant with p values between .01 and .05 which point the significant agreement between contextually similar users on criteria importance.



(a) TripAdvisor Data

(b) Educational Data

Figure 2: Following the rule of thumb, correlations close to +0.70 or -0.70 show a strong relationship; correlations closer to +0.5 and -0.5 indicate a moderate relationship; and correlations less than +0.5 and -0.5 indicate a weak relationship. Student test significance : p-value ≤ 0.05.

Consequently, we could conjecture that the more similar the users' contextual situations, the more these users tend to have similar criteria importance, as outlined by our hypothesis **H**.

Qualitative analysis. We further the above quantitative analysis with a qualitative analysis to get a better insight on the posed assumption **H**. We depict in Table 3 an illustration of users in different contextual situations that rated the same item from TripAdvisor dataset where our model leveraging criteria preferences is able to find the relevant hotels to recommend. We can see from Table 3 that the contextually similar users in the three first rows (u_1, u_2 and u_3) are interested in the same four item criteria (*value of the money, sleep quality, cleanliness and service*) among the 7 available criteria in the TripAdvisor dataset and they have close preferences on these criteria which make them very strongly correlated. In fact, for a summer family trip, it seems obvious to give importance for the value of the money criterion since hotels can cost a lot over summer per person and the hotel total expenses include the expenses of each family member. Besides, there could be extra costs per night for children. Furthermore, for a nice family vacation, it is important to be satisfied by the

Contextual situation	**User**	**Item**	**Rated Criteria**	**Overall Rating**
Summer 2011, traveled with family	u_1	Miramonti Hotel	value: 4; sleep quality: 4; cleanliness: 4; service: 5	4.2
Summer 2011, traveled with family	u_2	Miramonti Hotel	value: 4; sleep quality: 5; cleanliness: 5; service: 5	4.9
Summer 2011, traveled with family	u_3	Miramonti Hotel	value: 3; sleep quality: 5; cleanliness: 5 ; service: 4	4.8
Winter 2011, traveled as couple	u_4	Miramonti Hotel	location: 5; value: 4; Sleep Quality: 5; Rooms: 4; Cleanliness: 5; Service: 3	3.5
Autumn 2013, traveled as couple	u_5	Miramonti Hotel	location: 5; value: 4; Sleep Quality: 5 ; Rooms: 4; Cleanliness: 5; Service: 4	3.8

Table 3: Example of users preferences in different contextual situations from the TripAdvisor context-aware multi-criteria rating dataset

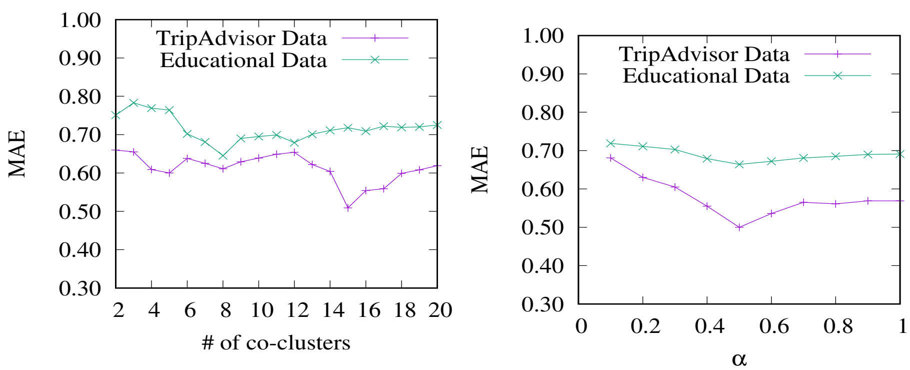
sleep quality, the delivered service and the cleanliness. For the users u_4 and u_5, we remark that they have different criteria preferences than u_1, u_2 and u_3. They are interested in other criteria such as location and room when selecting a hotel under different travel contexts such as traveling in the low season as couple. In fact, the location criterion plays a major role in choosing a hotel in such contextual situation, since it is important for a couple to choose a hotel with an incredible location surrounded by beautiful views and at the same time suitable for the low season climate. Added to that, it is of interest to have a quiet and comfortable room. By applying our predictive model, we consider the criteria preferences associated to each group of clustered contextually similar users to obtain criteria ratings for each co-cluster, where the first cluster includes the users u_1, u_2 and u_3, the second cluster includes u_4 and the last one includes u_5. Then, we obtain the overall ratings of the item "Miramonti Hotel" with higher values for u_1, u_2 and u_3 which will be recommended to these users.

For u_4 and u_5 which have different contextual situations than the first cluster users and thus belonging to other clusters, we notice that their different criteria preferences lead to have different overall ratings lower than those obtained for u_1, u_2 and u_3.

Consequently, these statistical experiments give us concrete proof of the assumption we make and provide a strong support for our research hypothesis **H**.

5.2.2. Parameter Tuning

We begin by conducting experiments of the proposed model on both TripAdvisor and Educational datasets with varying the co-cluster number L from 2 to 20. The impact of this parameter on the rating prediction accuracy results are plotted in Figure 3a. From this figure, we can observe that when the co-



(a) Impact of the co-clusters number

(b) Impact of the parameter α

Figure 3: The recommendation accuracy according to the co-clusters number and the parameter α

clusters number is equal to 8 the MAE value on the Educational dataset decline to the lowest. So, we come to a conclusion that 8 co-clusters is a better choice for our model on this dataset. While on TripAdvisor dataset, we note that our model requires 15 co-clusters to obtain maximum prediction accuracy. The co-clusters number when our model achieves the best prediction is reported for each dataset. Another important parameter that should be tuned is α (See formula (10)) to trade-off between the two bipartite graphs of our tripartite network. So we turned the value of α in the interval [0, 1] and we found that the prediction accuracy results changed along with the changing of α in Figure 3b. In fact, our model shows a poor accuracy when α= 0.1 and the accuracy tends to be steady when the α value is close to 1. This means that one bipartite graph is dominant which causes the tripartite partitioning failure because co-clustering actually degrade to be working on only one bipartite graph. When α is around 0.5, we have the best MAE measure. So, we set this value for both datasets.

5.2.3. Comparative performance evaluation (RQ2)

Table 4 presents the obtained results measured by the F-score metric on both TripAdvisor and Educational datasets, where "Improv" indicates the proposed model improvement comparing to each baseline model. We can see that in the two datasets, in the most of the cases, the recommendation models that consider the contextual information only (i.e., `CAMF`) or the multi-criteria ratings only (i.e., `ABM`, `CIC`, `CCA`) generally can achieve better results than the traditional model `BiasedMF` which doesn't consider neither contextual information nor multi-criteria preferences. For example, in the TripAdvisor dataset, the F-score value of `BiasedMF` is increased by +36.4% when integrating the contextual information using `CAMF` and by +46.2% when integrating the multi-criteria ratings using `CCA`.

When comparing with the recently proposed context-aware multi-criteria models (i.e., `DCL`, `CIL`, `CCIC`, `CDL`, `DCC`, `ICL`, `ICC`) which belong to the same category as our model, we can remark that in the TripAdvisor dataset, the F-score results are always greatly improved. The results in the Educational dataset show similar patterns for `CIL` and `CCIC` models. However, even when integrating the context to enhance the recommendation performance of the models based on dependent criteria rating prediction method (`DCL`, `CDL` and `DCC`), the F-score results decrease from 0.0720 for the traditional non contextual multi-criteria model `ABM` to 0.0660 and 0.0677 for `CDL` and `DCC` respectively. Besides, when integrating the contextual information into the second step only (i.e., the process of rating aggregations) using DCL, ICL and ICC, lower F-score results are seen in the Educational dataset.

		TripAdvisor dataset		*Educational dataset*	
Category	*Baselines*	*F-score*	*Improv*	*F-score*	*Improv*
Single-rating approach	*BiasedMF* [21]	0.0007	+67.1%	0.0698	+19.7%
Context-aware rating approach	*CAMF* [2]	0.0011	+48.4%	0.0715	+17.8%
Multi-criteria rating approaches	*ABM* [1]	0.0010	+53.1%	0.0720	+17.2%
	CIC [8]	0.0009	+55.4%	0.0700	+19.5%
	CCA [8]	0.0013	+38.9%	0.0682	+21.6%
	CCC [8]	0.0008	+60.5%	0.0644	+25.9%
Context-aware multi-criteria rating approaches	*DCL* [8]	0.0015	+29.6%	0.0620	+28.7%
	CIL [14]	0.0017	+20.2%	0.0749	+13.9%
	CCIC [14]	0.0018	+15.5%	0.0765	+12.1%
	CDL [14]	0.0015	+26.8%	0.0660	+24.1%
	DCC [14]	0.0019	+10.8%	0.0677	+22.2%
	ICL [14]	0.0016	+23.8%	0.0622	+22.1%
	ICC [14]	0.0017	+16.6%	0.0485	+44.3%
	Our model	**0.0021**	-	**0.0870**	-

Table 4: Comparison results on the TripAdvisor and Educational datasets

Regarding our model's results on the Educational dataset, it is shown that our model outperforms the comparative context-aware multi-criteria baselines in terms of F-score. Particularly, we can observe that it allows achieving an improvement of +13.9% and +12.1% over CIL and CCIC models which use independent contextual multi-criteria rating predictions and different rating aggregations methods. Thus, our solution advances the best performing baseline CCIC which uses an improved conditional aggregations. Comparing with the closest baseline CDL using dependent contextual multi-criteria rating predictions and linear aggregations, we could find an even more significant F-score improvement obtained by our model (+24.1%) on the Educational dataset. On the TripAdvisor dataset, our model also improves significantly the F-score over all the presented baselines models. More precisely, it even outperforms the closer baseline `CDL` by +26.8%. Regarding the `DCC` model that uses dependent contextual multi-criteria rating predictions and conditional aggregations, we find that the F-score value obtained by our model is slightly better than `DCC` by +10.8%.

5.2.4. Discussion

First, we design a research hypothesis based on the modeled entities and their relationships to provide insights about the desired co-clustering structure. To answer to RQ1, we conduct preliminary experiments aiming to determine the strength of the correlations between contextually similar users with respect to their criteria importance. As an answer to RQ1, we found a significant high correlation between users in similar contextual situations according to their criteria importance.

Turning now to the comparison between our co-clustering context-aware multi-criteria model and the context-aware multi-criteria baseline models, we conduct series of experiments. As an answer to RQ2, wa can draw several conclusions from the results obtained in Table 4. We begin by the comparison between the different baselines models included in the three first categories of recommendation approaches (single-rating approach, context-aware rating approach and multi-criteria rating approaches). We can notice an improvement that have been made by the context-aware and multi-criteria baselines models over the single rating model (`BiasedMF`). These results demonstrate that the extended recommendation approaches with context or criteria ratings information generally provide better results than traditional recommendation approaches when evaluating the top-N recommendation task using F-score metric. However, not all of the multi-criteria recommendation models (`CCA`, `CCC`) can obtain a better recommendation performance in terms of F-score in the Educational dataset. For instance, using the `CCA` model results in lower performance compared to the traditional model `BiasedMF` by -2.3%. Note that in the multi-criteria models, different methods are used for rating prediction (dependent or independent) which explain the performance variation between these models. The results reveal that the applied dependant method for providing criteria ratings by CCA and CCC may not be suitable for the Educational dataset. This may be because the nature of criteria, where students may present conflicting interests in the three criteria we have in this dataset.

With regard to the context-aware multi-criteria baseline models, a better F-score improvement can be achieved over the other baselines especially on the TripAdvisor dataset. On the Educational dataset, a better recommendation performance measured by F-score could be reached by contextualizing the multi-criteria ratings prediction step, but it undoubtedly depends on the applied criteria rating prediction method.

In comparison with the context-aware multi-criteria baseline models, we found that our proposed model outperforms all the presented baselines in terms of F-score, even the best performing baseline (CCIC) that employs independent contextual multi-criteria rating predictions and conditional aggregations. These results indicate that our model based on a dependant contextual multi-criteria rating predictions is able to beat the baselines using independent contextual multi-criteria rating predictions regardless of the used aggregation way. This points out that only improving the aggregation step fails to improve the recommendation results since it is important to improve the rating prediction on each criterion, as the final user's preference is estimated based on these criteria predicted ratings. Moreover, our model improves the closest baseline CDL using dependent contextual multi-criteria rating predictions and linear aggregations. This finding would seem to demonstrate that our proposed dependant method is valuable in solving the criteria ratings prediction problem, since it highlights not only the dependence between contexts and users in a low dimensional space but also emphasizes the correlations between criteria in the prediction process. Contrary to the dependent way used for criteria rating predictions by CDL that relies on incorporating and employing all the criteria ratings. In such case, when criteria are not in fact dependent, some useful information may be lost which can make the recommender less accurate with increasing its complexity. For instance, in the Educational dataset, the criterion "application" represents a student's taste on the domain of the projects, while "data" and "ease" criteria actually indicate the difficulty of the projects from the perspective of the students. Yet, some students want to select an easy project, while others may want to choose more challenging projects which results in some conflicting interests decreasing the performance when applying the criteria rating prediction dependent way. We have addressed this issue by considering only relevant criteria rated by contextually similar users in the same cluster which explain the superiority of our model in the Educational dataset.

The results on the TripAdvisor dataset show nearly similar patterns when evaluating the performance on top-N recommendations. However, the closeness of the results obtained by `DCC` model and ours may indicate that further work needs to be done on improving the aggregation step besides improving the first step in our context-aware multi-criteria model.

We conclude that the recommendation performance is improved due to the main delivered contributions that can transcend the previous recommendation approaches. In fact, the present article extends our previous work [28] that introduced a recommendation model based on bipartite graph partitioning. In the following, we summarize our contributions (including that of our previous research [28]) over previous work ([13, 14]):

(1) In our previous contribution [28], we proposed a novel approach of context-aware multi-criteria-based recommendation that explores the idea of clustering situational recommendations embedding contextually similar users evaluating items with respect to multiple facets. It primarily relies on the bipartite spectral graph co-clustering for jointly partitioning users' contextual situations and the rated items' criteria. The obtained co-clusters provide partial user's item ratings that are aggregated to predict the overall item rating using prioritized aggregation operators which allow tailoring the criteria strengths to the user's preferences.

(2) The key contributions of this extended work are presented as follows:

- Unlike what was previously used by the closest work [13, 14] to model the multi-dimensional available data, we present the relevant entities arising from linked heterogeneous recommendation data in the form of a tripartite graph including three types of connected entities (users, contextual situations and criteria). More specifically, we extend our prior work [28] based on the bipartite graph modeling to deal with users nodes and different context nodes including a set of contextual dimensions values representing the users' contextual situations. We also highlight a new challenge through the tripartite graph representation, consisting in weighting differently the three mentioned entities relationships on the one hand, and jointly clustering these entities on the other hand:
  - To weight the relatedness between users and contextual situations, we propose a frequency based weighting measure motivated by the TF-IDF (Term frequency inverse document frequency) scheme [15].
  - Based on the hypothesis that users in similar contextual situations tend to have similar interests for similar criteria, we explore the idea of triplet data co-clustering including contextually similar users providing similar criteria ratings to target items. Precisely, we replace the two-order co-clustering performed in our prior work by a high-order co-clustering modeled as the fusion of pair-wise sub problems over two bipartite graphs. For this purpose, we apply a collective factorization [17] to handle the tripartite graph partitioning problem based on the concept of consistency. The use of this type of co-clustering is still under-investigation especially in recommendation systems area.
- Unlike applying a traditional rating prediction algorithm as in our previous work [28], we explore a novel way to predict cluster-based multi-criteria ratings for users involved in similar contextual situations. Its main originality compared to the closest work [14], is its ability to consider the dependence between contexts and users in a low dimensional space using

a user splitting approach [18], and also emphasizes the correlation between criteria in the prediction process by an efficient correlation-based rating prediction algorithm [3].

- We perform an intensive comparative evaluation with state-of-the-art baselines belonging to four categories of work: single rating based methods, context-aware based rating methods, multi-criteria rating based methods and context-aware multi-criteria rating based methods including a very recent close work [14]. Then, we demonstrate the effectiveness of our approach over closely recommendation models.

Despite the extensive work reported in this article, there are challenges that need to be addressed for the future investigation. At the current stage our work only focuses on improving the first step in the aggregation-based multi-criteria recommendation by exploring a new way to predict the cluster-based multiple criteria. In the future, we plan to enhance the second step by designing an appropriate aggregation tied to user's preferences over the multiple item criteria resulting from the co-clusters. This line of work would give insight into the relevance of filtering pertinent item criteria before applying the aggregation. Furthermore, we plan to extend the proposed model by including more types of links between the graph entities. One example would be to additionally represent user-to-user links by considering social ties among users to take advantage of social influence. For improving the experimental evaluation, we plan to explore more data from the TripAdvisor dataset in order to get a larger dataset. For the Educational dataset, we plan to explore new solutions to better take advantage of the multi-criteria ratings to overcome the "conflicting interests" issue.

6. Conclusion

Personalized recommendations are an important part of a huge number of on-line applications related to diverse domains such as health (recommending healthy food), tourism (recommending sites to visit), and e-commerce (recommending products). Thus, a big challenge is to target the diversity of users in the one hand and the diversity of items to be recommended to those users in the other hand. Besides, it is well known today that recommender systems not only directly impact consumers' behaviour but also affect sales, revenue rates and business stakeholders. For increasing business value, it becomes increasingly crucial to better target the users' profiles as well as their dynamic changes with respect to the contextual situations of item purchase.

Strategically positioned at this target, we proposed a theoretical model that can support personalized recommendations at a fine-grained level description of users (through their contextual preferences) as well as items (through their criteria and values).

To build the predictive model, we first investigate how to model the multi-dimensional input data including users feedbacks on items criteria, and the contextual situations in which these users are involved when selecting items. Therefore, we present the context-aware multi-criteria network by interconnected multi-type entities in the form of a tripartite graph. Furthermore, we hypothesize that the users in similar contextual situations tend to provide similar assessments on the same item criteria, which motivates us to employ a high order co-clustering offering more personalized suggestions. Then, we exploit the obtained co-clusters with a new strategy for criteria ratings prediction which emphasizes the complementarities between the users and their contextual situations as well as between the different criteria. The obtained partial user's item ratings are finally aggregated to estimate the overall impression of an item.

As a direct implication, the predictive model proposed in this work can be exploited as a key building box of an interactive decision support system, which would increase the quality of recommendations. This is confirmed by our experiments undertaken on two real-world datasets. Those interactions could be put in the loop of the model training to endow the recommender system with self-learning abilities.

References

[1] G. Adomavicius, Y. Kwon, New recommendation techniques for multi-criteria rating systems, IEEE Intelligent Systems 22 (3) (2007) 48–55. doi:10.1109/MIS.2007.58.

[2] L. Baltrunas, B. Ludwig, F. Ricci, Matrix factorization techniques for context aware recommendation, in: Proceedings of the Fifth ACM Conference on Recommender Systems, RecSys '11, Association for Computing Machinery, New York, NY, USA, 2011, p. 301–304.
URL https://doi.org/10.1145/2043932.2043988

[3] Y. Zheng, B. Mobasher, R. Burke, Incorporating context correlation into context-aware matrix factorization, in: Proceedings of the 2015 International Conference on Constraints and Preferences for Configuration and Recommendation and Intelligent Techniques for Web Personalization - Volume 1440, CPCR+ITWP'15, CEUR-WS.org, 2015, p. 21–27.

[4] H. Wang, P. Li, Y. Liu, J. Shao, Towards real-time demand-aware sequential poi recommendation, Information Sciences 547 (2021) 482 – 497. doi:https://doi.org/10.1016/j.ins.2020.08.088.

[5] G. Adomavicius, R. Sankaranarayanan, S. Sen, A. Tuzhilin, Incorporating contextual information in recommender systems using a multidimensional approach, ACM Transactions on Information Systems (TOIS) 23 (1) (2005) 103–145.

[6] Z. Cheng, J. Shen, On effective location-aware music recommendation, ACM Transactions on Information Systems (TOIS) 34 (2) (2016) 1–32.

[7] S. Jiang, X. Qian, J. Shen, Y. Fu, T. Mei, Author topic model-based collaborative filtering for personalized poi recommendations, IEEE transactions on multimedia 17 (6) (2015) 907–918.

[8] Y. Zheng, Criteria chains: A novel multi-criteria recommendation approach, in: Proceedings of the 22nd International Conference on Intelligent

User Interfaces, IUI '17, Association for Computing Machinery, New York, NY, USA, 2017, p. 29–33.
URL https://doi.org/10.1145/3025171.3025215

[9] R. S. Sreepada, B. K. Patra, A. Hernando, Multi-criteria recommendations through preference learning, in: Proceedings of the Fourth ACM IKDD Conferences on Data Sciences, CODS '17, Association for Computing Machinery, New York, NY, USA, 2017.
URL https://doi.org/10.1145/3041823.3041824

[10] S. Gupta, V. Kant, An aggregation approach to multi-criteria recommender system using genetic programming, Evolving Systems 11 (1) (2020) 29–44. doi:10.1007/s12530-019-09296-3.

[11] S. Wang, J. Yang, Z. Chen, H. Yuan, J. Geng, Z. Hai, Global and local tensor factorization for multi-criteria recommender system, Patterns 1 (2) (2020) 100023. doi:https://doi.org/10.1016/j.patter.2020.100023.

[12] A. Rashed, J. Grabocka, L. Schmidt-Thieme, Attribute-aware non-linear co-embeddings of graph features, in: T. Bogers, A. Said, P. Brusilovsky, D. Tikk (Eds.), Proceedings of the 13th ACM Conference on Recommender Systems, RecSys 2019, Copenhagen, Denmark, September 16-20, 2019, ACM, 2019, pp. 314–321. doi:10.1145/3298689.3346999.
URL https://doi.org/10.1145/3298689.3346999

[13] Q. Li, C. Wang, G. Geng, Improving personalized services in mobile commerce by a novel multicriteria rating approach, in: Proceedings of the 17th International Conference on World Wide Web, WWW '08, Association for Computing Machinery, New York, NY, USA, 2008, p. 1235–1236.
URL https://doi.org/10.1145/1367497.1367743

[14] Y. Zheng, S. Shekhar, A. A. Jose, S. K. Rai, Integrating context-awareness and multi-criteria decision making in educational learning, in: Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, SAC

'19, Association for Computing Machinery, New York, NY, USA, 2019, p. 2453–2460.
URL https://doi.org/10.1145/3297280.3297522

[15] G. Salton, C. Buckley, Term-weighting approaches in automatic text retrieval, Information processing & management 24 (5) (1988) 513–523.
URL https://doi.org/10.1016/0306-4573(88)90021-0

[16] B. Gao, T.-Y. Liu, X. Zheng, Q.-S. Cheng, W.-Y. Ma, Consistent bipartite graph co-partitioning for star-structured high-order heterogeneous data co-clustering, in: Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, KDD '05, New York, NY, USA, 2005, pp. 41–50.
URL https://doi.org/10.1145/1081870.1081879

[17] A. P. Singh, G. J. Gordon, Relational learning via collective matrix factorization, in: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '08, Ny, USA, 2008, pp. 650–658.
URL https://doi.org/10.1145/1401890.1401969

[18] Y. Zheng, R. Burke, B. Mobasher, Splitting approaches for context-aware recommendation: An empirical study, in: Proceedings of the 29th Annual ACM Symposium on Applied Computing, SAC '14, ACM, New York, NY, USA, 2014, pp. 274–279.
URL https://doi.org/10.1145/2554850.2554989

[19] Y. Zheng, Educational group recommendations by learning group expectations, in: 2019 IEEE International Conference on Engineering, Technology and Education (TALE), 2019, pp. 1–7. doi:10.1109/TALE48000.2019.9225968.

[20] L. Baltrunas, F. Ricci, Context-based splitting of item ratings in collaborative filtering, in: Proceedings of the 3th ACM Conference on Rec-

ommender Systems, RecSys '09, NY, USA, 2009, pp. 245–248. doi:10.1145/1639714.1639759.

[21] Y. Koren, R. Bell, C. Volinsky, Matrix factorization techniques for recommender systems, Computer 42 (8) (2009) 30–37. doi:10.1109/MC.2009.263.

[22] S. Ruder, An overview of gradient descent optimization algorithms, arXiv preprint arXiv:1609.04747.

[23] M. Nilashi, D. Jannach, O. bin Ibrahim, N. Ithnin, Clustering-and regression-based multi-criteria collaborative filtering with incremental updates, Information Sciences 293 (2015) 235–250.

[24] D. Jannach, M. Zanker, M. Fuchs, Leveraging multi-criteria customer feedback for satisfaction analysis and improved recommendations, Information Technology Tourism 14 (2) (2014) 119–149.
URL https://doi.org/10.1007/s40558-014-0010-z

[25] S. Renjith, A. Sreekumar, M. Jathavedan, An extensive study on the evolution of context-aware personalized travel recommender systems, Information Processing Management 57 (1) (2020) 102078. doi:https://doi.org/10.1016/j.ipm.2019.102078.

[26] Y. Zheng, B. Mobasher, R. Burke, Carskit: A java-based context-aware recommendation engine, IEEE Computer Society, USA, 2015.
URL https://doi.org/10.1109/ICDMW.2015.222

[27] G. Denaro, S. Morasca, M. Pezzè, Deriving models of software fault-proneness, in: Proceedings of the 14th international conference on Software engineering and knowledge engineering, Vol. 27, 2002, pp. 361–368. doi:10.1145/568760.568824.

[28] R. Dridi, L. Tamine, Y. Slimani, Context-aware multi-criteria recommendation based on spectral graph partitioning, in: Database and Expert Systems

Applications - 30th International Conference, DEXA 2019, Linz, Austria, August 26-29, 2019, Proceedings, Part II, 2019, pp. 211–221.
URL https://doi.org/10.1007/978-3-030-27618-8_16