Multi-Criteria User Modeling in Recommender Systems

Kleanthi Lakiotaki[1], Nikolaos F. Matsatsinis[1]

and Alexis Tsoukiàs[2]

[1]Decision Support Systems Laboratory
Technical University of Crete
University Campus, 73100, Chania, Greece
{klio, nikos}@ergasya.tuc.gr
[2]LAMSADE – CNRS
Universite Paris Dauphine
Place du Marechal de Lattre de
Tassigny 75775, Paris ‑ France
Alexis.Tsoukias@dauphine.fr

Abstract

Recommender systems are software applications that attempt to reduce information overload. Their goal is to recommend items of interest to the end users based on their preferences. To achieve that, most Recommender Systems exploit the Collaborative Filtering approach. In parallel, Multiple Criteria Decision Analysis (MCDA) is a well established field of Decision Science that aims at analyzing and modeling decision maker's value system, in order to support him/her in the decision making process. In this work, a hybrid framework that incorporates techniques from the field of MCDA, together with the Collaborative Filtering approach, is analyzed. The proposed methodology improves the performance of simple Multi-rating Recommender Systems as a result of two main causes; the creation of groups of user profiles prior to the application of Collaborative Filtering algorithm and the fact that these profiles are the result of a user modeling process, which is based on individual user's value system and exploits Multiple Criteria Decision Analysis techniques. Experiments in real user data prove the aforementioned statement.

Keywords: Recommender Systems, Multiple Criteria Decision Analysis, Disaggregation – Aggregation approach, Collaborative Filtering, Preference Modeling, User Modeling and Clustering.

1. Introduction

A Recommender System (RS) will potentially suggest to the end user to watch or not a movie, to buy or not an item, to listen or not to a song and so forth. In this sense, an accurate Recommender System will ideally be able to act on behalf of the user [1]. To achieve its goal it must gain knowledge of user's value system and decision policy. Most existing recommender systems use the so called Collaborative Filtering approach (CF), some others are based on the content-based approach and many attempts combine these two methods into hybrid frameworks. Several reviews exist on Recommender Systems, one of the most representative being that of Adomavicius and Tuzhilin [2].

Personalization and customization are considered increasingly important elements of marketing applications. *"Personalization is the ability to provide content and services that are tailored to individuals based on knowledge about their preferences and behavior"* is just one of the various definitions attributed to the concept of personalization [3]. These terms usually refer to exploiting information about a user (that may be a customer, an individual, or a group) to better design products and services targeted to that user. Recommender Systems (RS), systems that assist users in discovering the closest to their preferences item, constitute an intrinsically important tool of personalization technologies. Nevertheless, the effectiveness of personalized services highly depends on user profile completeness and accuracy and thus, user modeling has become a key area in the development of personalization technologies.

So far, the majority of existing Recommender Systems obtains an overall numerical rating r_{ui}, as input information for the recommendation algorithm. This overall rating depends only on one single criterion that usually represents the overall preference of user u on item i. However, articles like [4] underline the pretence of stirring Recommender Systems researchers towards a more user oriented perspective, indicating that people are not truly satisfied by existing Recommender Systems.

The key to more effective personalization services is the ability to develop a system able to understand not only what people like, but why they like it. In other words, an accurate modeling of user value system and an effective preference representation schema, will potentially lead to the design of a recommendation algorithm with increased performance. Such a system can understand

how users think about items, by considering the knowledge about the underlying attributes that attract users to choose this particular item and hence recognize preferences, not just patterns, ensuring a more sophisticated understanding of the user.

In Decision Theory, the field of Multiple Criteria Decision Analysis (MCDA) emerged from the fact that real world decision making problems are intrinsically multidimensional [5]. MCDA aims at giving the decision maker a recommendation, in other words aiding the decision maker in the so called decision making process, concerning a set of objects, actions, alternatives, items etc, evaluated on multiple points of view, which are roughly referred as criteria (attributes, features, variables etc). In abstract, we may principally claim, that the word "aiding" in MCDA implies that the decision model representing decision maker's value system "supports" and in no case substitutes the decision maker during the decision process.

Under such a perspective, we face the recommendation process as a decision problem and exploit techniques from Decision Theory and more specifically from the field of MCDA, to accurately build a model representing the user's preferences. Following this framework, a potential user in a Recommender System corresponds to the decision maker in a Decision Process. We can thus consider Recommender Systems as Decision Support Systems where the Decision Maker's preferences need to be learned (with a necessary cognitive effort). Both, Decision Support Systems and Recommender Systems, try to assist the decision maker and user, respectively, throughout the decision making process. This decision may vary, from a simple purchase of an item, to more sophisticated managerial matters.

In the presented work, a hybrid methodological framework is proposed that combines techniques from the field of Multiple Criteria Decision Analysis and more specifically from the Disaggregation-Aggregation approach to model user's preferences, together with the Collaborative Filtering technique from the field of Recommender Systems, to identify the most preferred unknown items for every user.

2. Related work

Already in the 18th century, Condorcet (1743-1794), divided the decision process into three stages, the "first discussion phase", where the principles that would serve as the basis for the decision are discussed, the "second discussion phase", in which the question is clarified, opinions approach and

combine with each other to a small number of more general opinions and also the alternatives are determined and the "third phase", which consists of the actual choice between these alternatives. Later on, Simon [6] adjusted the existent approaches to become suitable for decisions in organizations into three phases, the "intelligence", the "design" and the "choice" phase. A more recent paper of Tsoukiàs [7], introduces a descriptive model of the "decision aiding process", that is the set of activities occurring between a Decision Maker and an Analyst who develops a formal model aiming at helping the Decision Maker to face a problem situation. This model considers the decision aiding process as a cognition process, introducing schematically the cognitive artifacts aiming at supporting the Decision Maker's decision process. Within such a model a *recommendation* results from the construction of an *evaluation model* resulting from a *problem formulation* which represents formally a specific *problem situation*.

The field of Multiple Criteria Decision Analysis (MCDA) is a well established field of Decision Science, and comes into a large variety of theories, methodologies, and techniques [5]. Multiple Criteria Decision Analysis aims at assisting a decision maker in dealing with the ubiquitous difficulties in seeking compromise or consensus between conflicting interests and goals, represented by the roughly referred as "multiple criteria". A common approach states that Multiple Criteria Decision Analysis is a methodology enabling the construction of a reliable and convincing model when several alternatives need to be assessed against multiple attributes under different problem statements (choosing, rejecting, ranking, classifying etc.).

User modeling (UM) is a cross-disciplinary research field that attempts to construct models of human behavior within a specific computer environment. Some approaches of modeling user preferences have already been applied to Recommender Systems and mainly adopt techniques and methodologies from the greater field of Artificial intelligence, Knowledge Engineering, or Data Mining, like ontological user profiling [8], or from the general field of Statistics [9]. Yet, new ideas and approaches of UM appear throughout literature, indicating that user modeling emerges as an important functional tool to enhance the performance of RS's [10]. However, to our knowledge, an approach like the one described herein, that constructs user profiles by exploiting preference information that is integrated into user value system and utilizing these profiles to identify recommendation patterns, is considered novel in the field of Recommender Systems.

At times, significant research has been undertaken towards the detection of cross links between the MCDA methodological frameworks and other scientific fields and disciplines, like Artificial Intelligence and Machine Learning in order to enhance the preference modeling capabilities of MCDA approaches and to improve their overall performance and efficiency. However, to incorporate multi-criteria ratings in an existing recommendation process or to design new recommendation techniques, careful consideration is necessary, to achieve maximum accuracy [11]. Recently, Zhang et al. [12] introduced multi-criteria rating in Recommender Systems, from a statistical machine learning perspective however.

The UTARec system [13], a predecessor of the proposed system, is an initial demonstration of the application of Multi-Criteria Decision Analysis techniques in Recommender Systems. However, UTARec constituted only an experimental proof of the multicriteria algorithm efficiency to predict real user ratings and served as a stepping stone for the integrated hybrid Multi-Criteria Recommender System presented herein.

3. Methodological Framework

In this section, we discuss the overall framework of the proposed approach together with the system's individual components. Figure 1 summarizes our system's overall process structure, and the following subsections outline the steps involved. All the details are analytically discussed throughout this section, while the case study (a movie recommender systems described in section 6), ensures a straightforward and complete presentation of the proposed methodology.
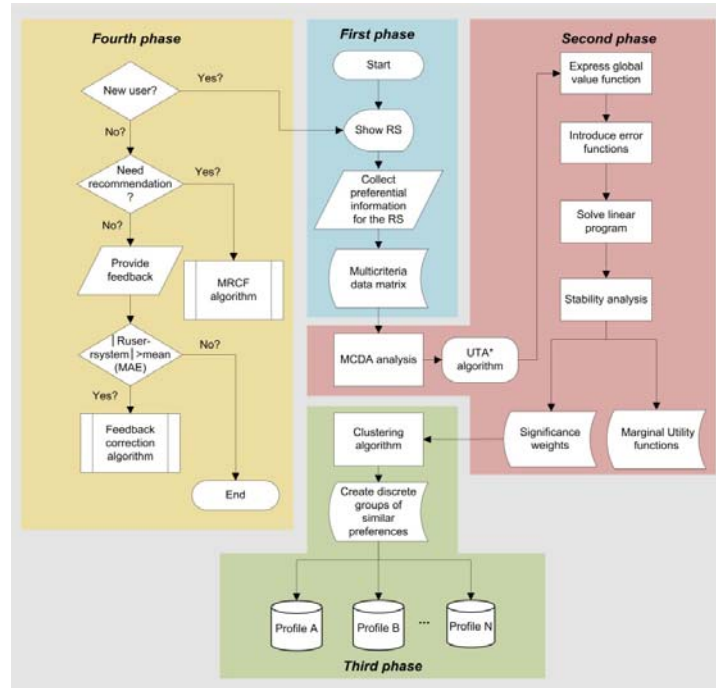
Figure 1: Proposed system's build up architecture

3.1 First phase: Data acquisition

Foremost to describing data acquisition procedure, it is essential to clarify at this point that two different types of data are gathered herein. Both types are attained by user statements. The first type concerns preference data given as numerical ratings, while the second type deals with preference statements in the form of a ranking order, the so called *weak preference order*.

To acquire user preference information, every user $u_t \in U$, where $t=1,2,...,n$, n being the total number of users, is asked to evaluate a set of items $A_i \in A_R$, named the reference set A_R. For every alternative $A_i \in A_R$, $i=1,2,...,m$, where m is the length of A_R, the user u provides a rating r_{ui}, for every criterion c_j, $j=1,2,...,k$, where k is the total number of criteria, following a predefined measurement scale (i.e. 1 to 5). Additionally to these individual evaluations, the user is asked to rank in a descending order all the alternatives that belong to the reference set and thus provide a *weak preference order*. Indifference relations are acceptable in the ranking order and are considered accordingly during the multicriteria user modeling phase. With the completion of the data

acquisition step, an *n* by *k+1* data matrix is formed that acts as an input for the second phase. An example of such a matrix is show in Table 2 of section 6.

3.2 Second phase: Multicriteria user modeling

The multi-criteria input data matrix as acquired from the first phase, is analyzed and processed throughout the system's second phase, leading to the formation of a single k-dimensional vector for every user, referred as the *significance weight vector* or merely the *weight vector*. During the multicriteria user modeling phase, the UTA* algorithm [14], one of the most representative and widely applied Disaggregation-Aggregation framework algorithms is applied, to analyze user's cognitive decision policy. The UTA* algorithm, adopts the preference disaggregation principle, the philosophy of which is to assess/infer preference models from given preferential structures. A complete step series of the Disaggregation-Aggregation approach is shown in Figure 2.
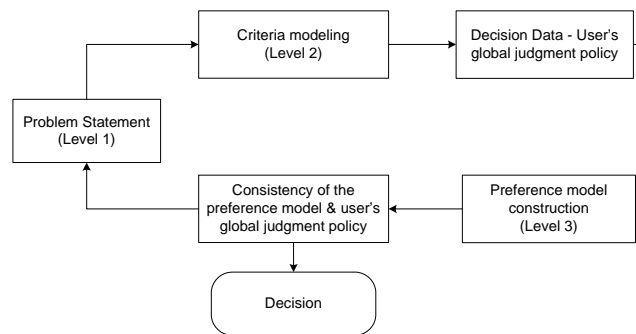

Figure 2: The Disaggregation – Aggregation approach

The first step of the Disaggregation-Aggregation approach deals with determining the problem statement (or problematic) in which the examining problem belongs to. Among the various problem statements that are met in Decision Aiding theory [15], three problem statements are mostly appropriate in the case of a *recommendation problem*. These are: *choosing* one or more potential action/s from a set of actions (alternatives) *A*, *ranking* those alternatives in a descending order, or *sorting* them into predefined ordered categories. In general, there are various ways to present recommendations to the end user; either by offering the user the best item (choosing), or by presenting the top N items as a recommendation list (ranking), or by classifying the items into categories, i.e. "highly recommended", "fairly recommended", "not recommended" (sorting).

Accordingly, a recommendation problem can equivalently belong to one of the first three problem statements, depending on its design architecture.

We need to clarify at this point that although the UTA method that is performed in this step through the UTA* algorithm, belongs to the *ranking* problem statement, this does not imply that the recommendation problem should also belong to the same problem statement. To elucidate the inconsistency that seems to emerge at this point, we simply mention that at the user modeling phase the problem to solve is to model user's value system and this is accomplished by means of the UTA method, however, the ultimate problem to solve is to predict ratings for unknown items.

Following the Disaggregation- Aggregation methodological schema, the modeling process of level 2 must conclude to a consistent family of criteria $\{g_1, g_2,...,g_k\}$. More details on the criterion family requirements can be also found in [5]. We briefly mention here that each criterion must be a non-decreasing, real valued function, defined on A, as follows:

$$g_j : A \rightarrow [g_{j*}, g_j^*] \subset \square \ / \ a \rightarrow \mathbf{g}(a) \in \square \qquad (1)$$

In 1, $[g_{j*}, g_j^*]$ is the criterion evaluation scale, g_{j*} and g_j^* are the worst and the best level of the jth criterion respectively, $g_j(\alpha)$ is the evaluation or performance of action α on the jth criterion and $\mathbf{g}(\alpha)$ is the vector of performances of action α on the k criteria.

The multicriteria data input matrix is processed by the UTA* algorithm through an iterative ordinal regression procedure. Analytical details and an illustrative example of the UTA* algorithm can be found in [14].

In abstract, the UTA* algorithm, considers as input a weak-order preference structure on a set of actions, together with the performances of the alternatives on all attributes, and returns as output a set of additive value functions based on multiple criteria, in such a way that the resulting structure would be as consistent as possible with the initial structure given by the user. This is accomplished by means of special linear programming techniques.

Four basic steps are followed in UTA* (see Appendix for more details) according to which, all the necessary parameters to estimate global value functions for each item and user are calculated. Thus,

a value is assessed for each alternative that belongs to the reference set, quantifying its value to each user and ensuring consistency with the user's value system. UTA*'s output involves the value functions associated to each criterion, approximated by linear segments, as well as the criteria significance weights (trade-offs among the criteria values). The later, expressed as a *weight vector* for every user, serves as his/ her value system information representation schema and provides the required user modeling data to proceed to the third phase, the clustering phase.

3.3 Third phase: Clustering

Generally, a clustering algorithm divides the original data set into disjointed groups. Clustering is an unsupervised process aiming at grouping data objects, based only on information found in the data that describes the objects and their relationships. The goal of a clustering algorithm is that the objects within a group should be similar (or related) as much as possible to one another, while they should be different from (or unrelated to) the objects in other groups. Most of existing clustering algorithms like the popular for its simplicity k-means are sensitive to initial parameters, such as the number of clusters and initial centroid positions. To limit these shortcomings, the global k-means [16], a deterministic approach of the traditional k-means clustering algorithm, is enrolled in the third phase. Global k-means does not depend on any initial parameter values and employs the k-means algorithm as a local search procedure. Instead of randomly selecting initial values for all cluster centers, this algorithm acts in an incremental way, by optimally adding one new cluster centre at each stage, the one that minimizes a certain clustering criterion.

Suppose we are given a data set $\{x_1, x_2,...x_n\}$, $x_n \in R^d$. The k-clustering problem aims at dividing this data set into k disjoint groups called clusters $C_1, C_2,...C_k$, by optimizing a certain clustering criterion. The most widely used clustering criterion, adopted also in this case, is the Sum of Squared Error (SSE) between each data point x_i and the centroid m_j of the subset C_j which contains x_i. This clustering criterion depends on the cluster centers $m_1, m_2, ..m_k$, and is shown in equation 2.

$$SSE(m_1, m_2,...,m_K) = \sum_{i=1}^{N}\sum_{j=1}^{K} I(x_i \in C_j)|x_i - m_j|^2 \tag{2}$$

Global k-means is applied to the set of user *weight vectors* and labels every user to a specific group (user profile).

3.4 Fourth step: Recommendation phase

Following the formation of user groups with similar preferences (user profile clusters), accurate item recommendations can be provided to these users. The recommendation phase is accomplished by implementing the collaborative filtering philosophy inside each user group.

The *multidimensional Multi-criteria Collaborative Filtering (MRCF-dim)* approach applied herein, is based on multidimensional distance metrics. First, it calculates the distance between two users, u and u', for the same item, according to equation 3:

$$d_{uu'} = \sqrt{\sum_{n=1}^{k} \left(r_{un} - r_{u'n} \right)^2} \qquad (3)$$

In 3, r_u is the rating vector of user u and $r_{u'}$ the rating vector of user u'. By rating vector we mean the set of ratings that user u provided for an item i, including the overall rating and k represents the dimension of this rating vector, as a result of the criteria number and the overall rating that altogether define user vector's dimensionality.

Second, the overall distance between two users u and u' is simply given by equation 4.

$$dist(u,u') = \frac{1}{|U(u,u')|} \sum_{i \in U(u,u')} d_{uu'} \qquad (4)$$

In 4, *U(u,u')* denotes the set of items that both u and u' have rated. This means that the overall distance between two users, u and u', is the average distance between their ratings for all their common items.

Finally, users' similarity, which is inversely related to their distance, is given by:

$$sim(u,u') = \frac{1}{1 + dist(u,u')} \qquad (5)$$

This notion of similarity ensures that the similarity will approach 0 as the distance between two users becomes larger, and it will be 1 if two users rated all their common items evenly [11].

Note that *sim(u,u')*, is calculated if and only if, *u'* belongs to the same group with *u* and their *U(u,u')* is not empty (henceforth, we will refer to these users as "mates"). Therefore, the computational effort is minimized compared to traditional non clustering approaches that compute *sim(u,u')* for all possible user combinations.

After calculating a similarity index for "mate" users, equation 6 provides a potential rating *R(u,i)* for any unexplored item *i*.

$$R(u,i) = \left(\frac{1}{\sum_{u' \in C(u)} sim(u,u')} \right) \cdot \sum_{u' \in C(u)} sim(u,u') \cdot R(u',i) \qquad (6)$$

Equation 6 is in fact a similarity weighted sum of known ratings and *C(u)* defines user's neighborhood, meaning the cluster where *u* belongs to.

3.5 Feedback mechanism

System's feedback mechanism is activated by a user, when he/she is willing to provide a rating for an item that explored, according to system's preceding recommendations. In the case where a user disagrees with the recommendation given and provides the rating that he/she would give to the specific item, the system processes this information, by triggering the feedback correction algorithm. According to this algorithm, the new user value is compared to past system value in terms of absolute difference. If this difference is greater than the mean absolute difference stored for this user, then this alternative is included in the reference set and the UTA* algorithm runs again to calculate a new significance weight vector for this user. The weight vector will indicate whether this particular user should belong to a different group. To decide this, the feedback correction algorithm calculates the Squared Euclidean distance (SE) of the user's weight vector from every centroid of the formed groups. This particular user will now belong to the group where user's SE is less or equal to group's maximum SE. According to the results of the feedback function, the system updates, or not, the specific user profile, which in other words means, that may change or not the group that this user belongs to.

The aforementioned approach can be considered rather simplified, whereas more sophisticated alternatives (e.g. relevance feedback techniques) are investigated to be integrated into the system.

4. Other traditional Single and Multicriteria Rating Collaborative Filtering approaches

The contribution of our work can be stated as follows: **a)** we argue that multiple criteria **should** be considered to better understand user decision policy, with the purpose of constructing a decision model and be able to recommend items of interest by exploiting this model, **b)** we claim that the user modeling phase can be more effective if more sophisticated methods, specially developed to treat multiple criteria decisions, are used to build user profiles. It is important to understand *how* the user came up to a decision and not only consider his/ her past actions or other people's similar decision. It is very likely that two users that gave the same overall grade to an item, to passed through much unlike decision routes to reach at the same point. Finally **c),** we advocate that different users also have different knowledge, interests, abilities, learning styles, and preferences; this however, does not preclude the existence of discrete patterns among users, the so called user profile groups. As a consequence, in order to verify the effectiveness of our approach we compare it to popular collaborative filtering approaches that are analytically explained in the related literature [11]. More specifically, the proposed approach is compared to: **a)** a *single rating collaborative filtering approach (SR-CF)* that uses the *item weighted cosine-based similarity* to calculate similarities, **b)** a *multiple rating collaborative filtering* technique that uses *average similarity* to aggregate *item weighted cosine based similarities* from individual criteria *(MRCF-av),* and c) a MRCF technique that uses *worst-case similarity* to aggregate *item weighted cosine based similarities* from individual criteria *(MRCF-min)*.

4.1 Single rating collaborative filtering approach (SR-CF)

By single-rating we mean that just the overall rating is considered in all calculations. In *SR-CF* approach, a similarity index *sim(u,u')* is calculated for all possible *u-u'* combinations according to the *item weighted cosine similarity function* that uses the common notion of cosine similarity. However, even though cosine similarity measure has been extensively used in Recommender Systems, it fails to compute a rating in the case of a single common item. If the number of common items is 1, cosine similarity will result in 1, regardless of differences in individual ratings. Furthermore since cosine similarity does not consider the size of *U(u,u'),* we used a weighted approach of this measure as given by 7.

$$sim(u,u') = A \cdot \left(\frac{\displaystyle\sum_{i \in U(u,u')} R(u,i) \cdot R(u',i)}{\sqrt{\displaystyle\sum_{i \in U(u,u')} R(u,i)^2} \cdot \sqrt{\displaystyle\sum_{i \in U(u,u')} R(u',i)^2}} \right) \qquad (7)$$

$$A = \frac{U(u,u')}{U(u)} \qquad (8)$$

In 7, the traditional cosine similarity measure inside the parenthesis, is multiplied by *A*, which is the percentage of common items *U(u,u')* that both *u* and *u'* have rated, over *U(u)*, the total number of items that *u* has rated. Obviously, similarities of pairs of users with large common item set are favored.

4.2 Multi-rating collaborative filtering approaches (MRCF)

Two basic approaches have been proposed that include multi-rating information in similarity calculations [11]. The first considers individual similarities on different attributes, which are traditionally calculated by cosine similarity metrics and the second calculates similarities based on multidimensional distance metrics. The latter, is also adapted in the recommendation phase of the proposed methodology.

Following the first approach, various techniques are employed to aggregate individual similarities. We used two different aggregation ways, the *average similarity* and the *worst-case similarity.* Both calculate cosine similarities on all criteria as well as on the overall values. Their only difference is that in the *average similarity* approach these individual similarities are averaged, while in the *worst-case similarity* scenario the minimum of all attribute and overall similarities is chosen to represent users' similarity. However, even though we employed these two approaches as also mentioned in [11], cosine similarities in this work use the *item weighted* variation of cosine similarity as given in equation 7.

5. Assessment

In general, Recommender Systems have been evaluated in many, often incomparable, ways. A review on the key decisions in evaluating collaborative filtering recommender systems can be found in [17]. In this work we focus on the ways in which prediction quality is measured. To assess prediction quality, three different kinds of metrics were employed.

5.1 Statistical Accuracy Metrics

Statistical accuracy metrics measure how close is the numerical value r'_{ui}, which is generated by the Recommender System and represents the expected rating of user u on item i, to the actual numerical rating r_{ui}, as provided by the same user for the same item. The most commonly used statistical accuracy metric is the Mean Absolute Error (MAE). Since Mean Absolute Error measures the deviation of predictions generated by the Recommender System from the true rating values, as they were specified by the user, it is measured only for those items, for which user u has expressed his opinion. Suppose n is the number of items that user u has expressed an opinion, then, the MAE_u is formally given by equation 9.

$$MAE_u = \frac{1}{n}\sum_{i=1}^{n}\left|r_{ui} - r'_{ui}\right| \tag{9}$$

The average MAE for an entire data set, can be calculated by averaging the Mean Absolute Errors of all users, MAE_u, for $u = 1, 2, ...m$, over the total number of available users m and can give as an overall estimation of a model's performance.

Another very popular statistical accuracy metric is the Root Mean Squared Error. The difference in RMSE and MAE is that in MAE all the individual differences are weighted equally in the average, while in RMSE since the errors are squared before averaged, relatively high weight is given to large errors. The equation to calculate RMSE is given as:

$$RMSE_u = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(r_{ui} - r'_{ui}\right)^2} \tag{10}$$

Alike average MAE, average RMSE offer a global estimation of a model's prediction accuracy.

5.2 Classification accuracy metrics

Classification accuracy metrics determine the success of a prediction algorithm in correctly classifying items. In Recommender Systems, a rational classification of items would be as *"highly recommended"* and *"not recommended"*. Items of the first class are very likely to be proposed by the system, while items that belong to the second category will be never shown to the user.

Precision, the number of true positives, is the number of items correctly labeled as belonging to the *"highly recommended"* class, divided by the total number of items labeled as belonging to the same class. Recall is defined as the number of true positives divided by the total number of elements that actually belong to the *"highly recommended"* class. Since there is a trade-off between precision and recall, F-measure, a harmonic mean that equally weights precision and recall is often used.

5.3 Rank correlation coefficient

Kendall's tau is a measure of correlation between two ordinal-level variables. In order to calculate Kendall's tau for any sample of *n* items, there are *[n (n-1)/2]* possible comparisons of points (x_i, y_i) and (x_j, y_j). Suppose M_C is the number of pairs that are concordant, M_D is the number of discordant pairs and M is the total number of pairs. By concordant pair we mean that for the specific pair of items, both the user and the model ranked them identically. The formula for Kendall tau τ is:

$$\tau = \frac{M_C - M_D}{\sqrt{(M - I_Y) - (M - I_{\hat{Y}})}} \tag{11}$$

In 11, I_Y is the number of equivalent pairs regarding ranking order Y (user's ranking order) and $I_{\hat{Y}}$, is the number of equivalent pairs regarding ranking order $\hat{Y}$ (model's ranking order). Kendall's tau varies between -1 and 1 with 1 indicating a total agreement of the orders.

6. A multi-criteria movie recommender system

As an empirical study we applied the four phases of our proposed methodology to a movie Recommender System as described hereupon.

6.1 Data set description

In the commercial sector, Yahoo!Movies has launched a recommendation service that employs user-specific multi-criteria ratings for different movies (movies.yahoo.com). The experimental data

set was retrieved from Yahoo!Movies, where users provided preference information on movies based on four different criteria. The four attributes that constituted the criteria family were: c_1=*acting*, c_2=*story*, c_3=*direction* and c_4=*visuals*. All values were measured in a 13-fold qualitative scale with F denoting the worst evaluation grade and A+ declaring the most preferred value. For processing purposes, we replaced letters with numbers, so as 1 corresponded to the worst value, formerly denoted as F and 13 to best value, A+. In addition to individual criteria ratings, users were asked to provide an overall grade, which reflected their global preference over each movie. An example of the initial and the final, after our transformations rating scheme, is shown in Table 2. The left side shows a typical raw data form, while in the right side the same data is presented in a ready to process form (final data form). Data cleaning followed soon after the data acquisition phase to remove any case with at least one or more not available values. This introduced 18% shrinkage in our data set. A subsequent filter to cut off users with less than five rated movies was applied, to assure an adequate set of evaluated movies for every user. To this end, the resulting experimental data set included 6078 different users and 976 different movies. The total number of ratings was 62156 and every user had rated about 10 movies on average. The average evaluation grade was 9.6, 9.9, 9.5, 10.5 and 9.6 for the criteria *acting, story*, *direction*, *visuals* and *overall*, respectively. The Parson Correlation matrix of the data set is shown in Table 1.

	Story	Acting	Direction	Visuals	Overall
Story	1	0.834	0.871	0.782	0.905
Acting	0.834	1	0.857	0.786	0.865
Direction	0.871	0.857	1	0.835	0.911
Visuals	0.782	0.786	0.835	1	0.834
Overall	0.905	0.865	0.911	0.834	1
Correlation is significant at the 0.01 level					

Table 1: Pearson Correlation matrix of the data set.

It is crucial to mention at this point, that although recommender systems research society is used to very large data sets, i.e. the Netflix prize dataset consists of about 480 thousand users, these datasets consist of an overall single rating (usually in a scale of 1 to five stars) and do not provide any information on individual criteria. Generally, it is not very easy to come across data sets with preference information on several attributes, since it is commonly believed that people are unwilling to provide a lot of information. We advocate in this work that preference information on individual criteria offer valuable knowledge for the design and effectiveness of recommender systems, as it can be processed to build user's value system and decision policy, so, asking the user to provide this

information lead to significant improvement of recommendation accuracy, which can anticipate user's additional effort.

According to the methodological requirements of the Disaggregation-Aggregation approach as discussed earlier, a weak preference order of the alternatives is required to apply ordinal regression. The user provided that information, together with the performances on all four criteria for every movie of the reference set. However, because the global preference was expressed in a qualitative scale from 1 to 13, we transformed all global preference vales into a weak preference order for every user. For example, a sequence of numerical values like r_i = [13, 12, 12, 6, 1] when transformed into a ranking order will appear like, r'_i = [1, 2, 2, 3, 4].

Eventually, the multi-criteria data matrix, which acted as an input for the UTA* algorithm, consisted of the actual user ratings on all four criteria for the items belonging to the reference set, as well as of a weak preference order for these items. An example of the input multicriteria matrix is presented in the right side of Table 2.

Initial data form							Final data form						
user_id	Overall grade	C_1	C_2	C_3	C_4	movie_id	user_id	Ranking order	C_1	C_2	C_3	C_4	movie_id
1	A+	A+	A	A+	A-	1	1	1	13	12	13	11	1
	B+	B+	A+	B	A+	4		2	10	13	9	13	4
	B	B	A-	B	A+	25		3	9	11	9	13	25
	B-	B+	B+	B	B	23		4	10	10	9	9	23
	C+	C	B	C+	A+	9		5	6	9	7	13	9
2	A	A+	A-	A-	A+	9	2	1	13	11	11	13	9
	B+	B+	B	B	B	18		2	10	9	9	9	18
	B+	A-	A-	A+	B	2		2	11	11	13	9	2
	...	...	...	...	...	…		...	...	...	...	...	…

Table 2: A sample of the multicriteria data input matrix before (left side) and after (right side) preparation.

6.2 User modeling phase

The UTA* algorithm [14] processed the multicriteria data matrix to calculate significance weight vectors w_u ,for every user u. A matrix of 6078 × 4 was formed, which included the weight vectors of all users. All weights were normalized to a range from 0-1.

6.3 Clustering phase

Global k-means algorithm divided the 6078 weight vectors, resulted from the user modeling phase, into separate clusters. As already stated, global k-means ensures optimality at each clustering step. This means that SSE will continuously decrease over the number of clusters. A plot of SSE for different number of clusters is shown in Figure 3.
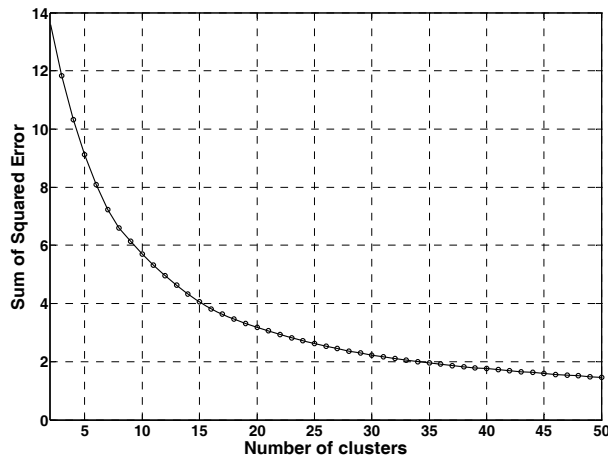

Figure 3: Sum of squared errors (SSE) versus the number of clusters

Although we can get a rough estimation of the clustering tendency from Figure 3, since the SSE will be constantly decreasing with the number of clusters, further investigation to identify the "optimal" number of clusters is necessary and most of times is application dependent. In the specific movie Recommender System application, the Mean Absolute Error MAE_u of every user, averaged for all clusters and plotted against the number of clusters, can provide a rough estimation of the "optimal" number of clusters. Even though this plot will continuously decrease over the number of clusters, we observe in Figure 4, that the curve's slope decreases as the number of clusters exceeds 20. Depending on the application however, any other meaningful metric or feature, or even various combinations of them, can be proved insightful for deciding how many group to keep.
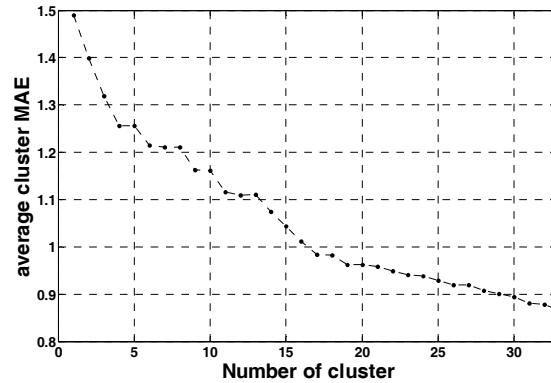
Figure 4: Average user Mean Absolute Error versus the number of clusters

The final outcome of the third phase is a collection of disjoint groups of users with similar preferences. These groups constitute the user profile clusters that the system's final step exploits to provide item recommendations. These groups can be updated when required as explained in 3.5.

6.4 Recommendation phase

During the user modeling phase a reference set of five movies constituted the *training set A_R* of each user. The residual rated movies were used as a *test set A_T*. Thus, during the recommendation phase a rating *R(u,i)* for every "unseen movie" $i \in A_T$ of user *u* was calculated according to equations 3-6.

7. Experimental results: Evaluation analysis

To demonstrate the accuracy and efficiency of the presented methodological framework as analytically described throughout section 3, we compared it to several traditional collaborative filtering techniques, as these are described in section 4 and calculated the two statistical accuracy metrics, MAE and RMSE, as well as precision, recall and Kendall's tau, in every case.

To effectively apply precision and recall measures, we divided the rating scale into two classes. The first, the *"highly recommended"* class, included only ratings ranged from 11 to 13, while all residual ratings belonged to the second class, the so called *"not recommended"* class. Notice here, that 42% of the overall preference values in our test set fall into the *"highly recommended"* class and so a precision of 42% corresponds to the threshold of a random guess.

To apply Kendall's tau, the values that our model predicted for all "unexplored" items *(test set)* of *u,* were transformed into a user's ranking order, r_u. The same logic was applied to the real ratings of this user and the two ranking orders were compared by applying equation 11.

Initially, we encountered the entire experimental data set as one group and applied all different collaborative filtering approaches. Table 3 shows the results of these approaches, first applied to ungrouped data and then to two different stages of our user modeling procedure, after 30 and 50 user profile clusters were formed. *SR-CF* corresponds to a single rating approach discussed earlier, and *MRCF-min* and *MRCF-ave* to the two multi rating collaborative filtering techniques of section 4.2. *MRCF-dim* corresponds to the multidimensional collaborative filtering technique that is already included and described in the proposed methodology. All predicted values emerge by applying equation 7 in different notions of similarity, depending on the method used. All similarity *sim(u,u')* and potential rating *R(u,i)* calculations, were implement in the *training set* of 6078 users, while evaluation metrics were calculated over the *test set* of the same users.

	MAE	RMSE	Precision	Recall	F-measure	Kendall's τ	
SR-CF	*2.464*	*2.733*	0.8341	*0.283*	*0,423*	0.865	*Ungrouped data*
MR-CF-min	2.358	2.620	*0.829*	0.337	0,479	0.869	
MR-CF-ave	2.430	2.695	0.830	0.299	0,440	0.867	
MR-CF-dim	**1.490**	**1.724**	**0.9163**	**0.7416**	**0,820**	**0.876**	
SR-CF	*1.772*	*2.103*	0.882	*0.537*	*0,667*	0.870	*30 user profile clusters*
MR-CF-min	1.696	1.992	*0.880*	0.557	0,682	0.875	
MR-CF-ave	1.761	2.077	*0.880*	0.540	0,669	0.873	
MR-CF-dim	**0.819**	**0.956**	**0.964**	**0.843**	**0,899**	**0.889**	
SR-CF	*1.451*	*1.798*	0.905	*0.630*	*0,743*	0.878	*50 user profile clusters*
MR-CF-min	1.380	1.700	0.905	0.645	0,753	0.881	
MR-CF-ave	1.439	1.779	*0.904*	0.633	0,745	0.880	
MR-CF-dim	**0.637**	**0.761**	**0.974**	**0.882**	**0,926**	**0.896**	

Table 3: Evaluation results of the single and multiple collaborative filtering approaches as applied to ungrouped users and when 30and 50 profiles are formed.

In Table 3 with **bold** we signed the best performances of all approaches for every metric, while worst values are denoted in *italic*. Moreover, the part of Table 3 that corresponds to ungrouped data, serves as a baseline to compare the performance of all methods on different clustering schemes.

We easily observe from Table 3, that among all methods, *MR-CF-dim,* which is the collaborative filtering method that is adapted in the proposed system, outperforms all other methods. It is also evident from Table 3, that user profile clusters improve the performance of any collaborative filtering method.

In Figure 5, we have plotted the average per user RMSE for an indicative part of users (700-750). Bars correspond to the RMSE values as a result of the proposed movie recommender system (when 30 clusters are formed), while the bullets correspond to RMSE values for ungrouped data which in turn corresponds to a traditional Multidimensional Collaborative Filtering System.
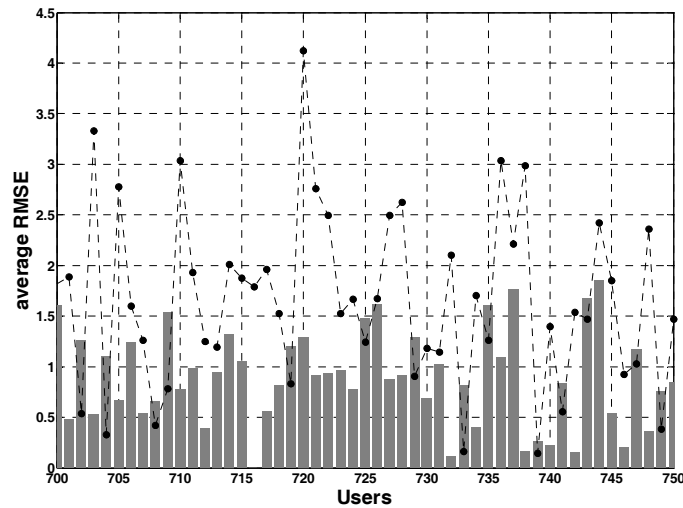

Figure 5: Average RMSE for 50 randomly selected users. Grey bars correspond to the average per user RMSE values when 30 user profile clusters are formed and bullets denote the equivalent values for ungrouped data.

Depending on the depth of personalization that each application poses, our system provides flexibility to examine every user individually. In Figure 6, for example, we plotted three evaluation metrics, MAE, Precision and Kendall's tau for a random user, vs. the number of clusters. We notice that for the specific user one rational decision would be to retrieve prediction values when 30 clusters are formed, where MAE has significantly decreased and Precision and Kendall' tau have almost reached their maximum levels. However, this decision may vary among users.
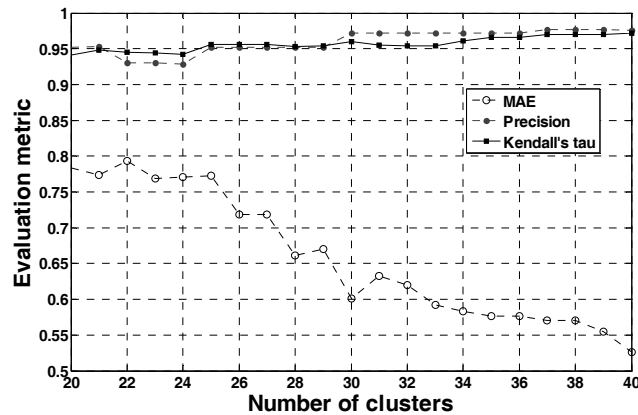
Figure 6: MAE, Precision and Kendall's tau vs. number of clusters for a random user.

For the specific user, if we consider a 30 user profiling scheme, the relative improvement in MAE, Precision and Kendall's tau, compared to the ungrouped case, is 41%, 16% and 3%, respectively. The differences are attributed to the diverse nature of each measure and may vary across users.

8. Conclusions

A multicriteria user modeling approach for recommender systems was presented in this paper. This approach was analytically explained through a hybrid methodological framework that incorporates techniques both from the field of Multicriteria Decision Analysis to model user preferences and from the field of Collaborative Filtering analysis to successfully recommend items to the end users. The proposed methodology was demonstrated as a movie Recommender System and its performance was tested in real user data. Furthermore, comparison study with other single and multiple criteria collaborative filtering methodologies clearly proves that the creation of user profiles constitutes an integral part of a recommendation process. These profiles are formed by an especially designed to treat multiple criteria approach, named the Disaggregation-Aggregation approach, as demonstrated by the UTA* algorithm. A sophisticated MCDA method, like the aforementioned, applied to build user profiles and the subsequent clustering of these profiles enhances the performance of any collaborative filtering technique used to predict preferences and offers new insights for Recommender Systems researchers to penetrate into novel, multidisciplinary, user modeling methodologies.

9. References

[1] Felfernig, A., Friedrich, G. and Schmidt-Thieme, L. 2007. Guest Editors' Introduction: Recommender Systems. IEEE Intelligent Systems. 22, 3, 18-21.

[2] Adomavicius, G. and Tuzhilin, A. 2005. Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE Transactions on Knowledge and Data Engineering. 17, 6, 734-749.

[3] Adomavicius, G. and Tuzhilin, A. 2005. Personalization technologies: a process-oriented perspective. Commun. ACM. 48, 10, 83-90.

[4] Zaslow, J. 2002. If TiVo Thinks You Are Gay, Here's How To Set It Straight The Wall Street Journal, sect. A (November 26,), 1.

[5] Figueira, J., Greco, S. and Ehrgott, M. *Multiple criteria decision analysis: State of the art surveys*. Springer, City, 2005.

[6] Simon, H. A. 1977 The New Science of Management Decision. Prentice Hall PTR.

[7] Tsoukiàs, A. 2007. On the concept of decision aiding process: an operational perspective. Annals of Operations Research. 154, 1, 3-27.

[8] Middleton, S. E., Shadbolt, N. R. and Roure, D. C. D. 2004. Ontological user profiling in recommender systems. ACM Transactions on Information Systems. 22, 1, 54-88.

[9] Zukerman, I. and Albrecht, D. W. 2001. Predictive Statistical Models for User Modeling. User Modeling and User-Adapted Interaction. 11, 1-2.

[10] Berkovsky, S., Kuflik, T. and Ricci, F. 2008. Mediation of user models for enhanced personalization in recommender systems. User Modeling and User-Adapted Interaction. 18, 3, 245–286.

[11] Adomavicius, G. and Kwon, Y. O. 2007. New Recommendation Techniques for Multicriteria Rating Systems. IEEE Intelligent Systems. 22, 3, 48-55.

[12] Zhang, Y., Zhuang, Y., Wu, J. and Zhang, L. 2009. Applying probabilistic latent semantic analysis to multi-criteria recommender system. AI Communications. 22, 2, 97–107.

[13] Lakiotaki, K., Tsafarakis, S. and Matsatsinis, N. *UTA-Rec: a recommender system based on multiple criteria analysis*. ACM, City, 2008.

[14] Siskos, Y., Grigoroudis, E. and Matsatsinis, N. 2005.UTA Methods. In Multiple criteria decision analysis:State of the art surveys, Springer, Boston, 297-344.

[15] Bouyssou, D., Marchant, T., Pirlot, M., Tsoukiàs, A. and Vincke, P. 2007 Evaluation and Decision Models with Multiple Criteria: Stepping stones for the analyst Springer.

[16] Likas, A., Vlassis, N. and Verbeek, J. 2003. The global k-means algorithm. Pattern Recognition. 36, 2, 451-461.

[17] Herlocker, J. L., Konstan, J. A., Terveen, L. G. and Riedl, J. T. 2004. Evaluating collaborative filtering recommender systems. ACM Trans. Inf. Syst. 22, 1, 5-53.

Appendix

The UTA* algorithm aims at estimating additive utilities of the form:

$$U(\mathbf{g}) = \sum_{i=1}^{m} u_i(g_i) \tag{1}$$

subject to the following constrains:

$$u_i(g_{i*}) = 0 \quad \forall i$$

$$\sum_{i=1}^{m} u_i(g_i^*) = u_1(g_1^*) + u_2(g_2^*) + \ldots + u_m(g_m^*) = 1 \tag{2}$$

where $u_i(g_i)$ $i=1,\ldots,m$ are non decreasing real valued functions, named marginal utility functions.

The UTA* algorithm may be summarized in the following steps:

Step 1: Express the global value of reference actions $u[g(\alpha_k)]$, $k = 1,2,\ldots,m$, first in terms of marginal values $u_i(g_i)$, and then in terms of variables w_{ij} according to the formula (3). The transformation of the global value of reference actions into weights values expression is made according to formula (4):

$$w_{ij} = u_i(g_i^{j+1}) - u_i(g_i^j) \geq 0, \quad \forall i = 1,2,\ldots,n$$
$$and \ j = 1,2,\ldots,a_i-1 \tag{3}$$

$$\begin{cases} u_i(g_i^1) = 0 & \forall i = 1,2,\ldots,n \\ u_i(g_i^j) = \sum_{t=1}^{j-1} w_{it} & \forall i = 1,2,\ldots,n \ and \ j = 2,3,\ldots,a_{i-1} \end{cases} \tag{4}$$

Step 2: Introduce two error functions σ^+ and σ^- on A_{Ri} (reference set of alternatives) by writing for each pair of successive actions in the given ranking the formula (5):

$$\Delta(a_k, a_{k+1}) = u[\mathbf{g}(a_k)] - \sigma^+(\alpha_\kappa) + \sigma^-(\alpha_\kappa)$$
$$- u[\mathbf{g}(a_{\kappa+1})] + \sigma^+(\alpha_{\kappa+1}) - \sigma^-(\alpha_{\kappa+1}) \tag{5}$$

Step 3: Solve the linear program (LP):

$$\left\{ \begin{array}{l} [\min]z = \sum_{\kappa=1}^{\mu}[\sigma^{+}(\alpha_k)+\sigma^{-}(\alpha_k)] \\[2mm] \text{subject to} \\[1mm] \left.\begin{array}{ll} \Delta(a_k,a_{k+1}) \geq \delta & \text{if } a_k \succ a_{k+1} \\[1mm] \Delta(a_k,a_{k+1}) = 0 & \text{if } a_k \,\square\, a_{k+1} \end{array}\right\} \ \forall k \\[3mm] \sum_{i=1}^{n}\sum_{j=1}^{a_i-1} w_{ij} \\[3mm] w_{ij} \geq 0, \ \sigma^{+}(\alpha_k) \geq 0, \ \sigma^{-}(\alpha_k) \geq 0 \ \ \forall i,j \text{ and } k \end{array}\right. \tag{6}$$

Step 4 (stability analysis): Check the existence of multiple or near optimal solutions of the linear program (6). In case of non uniqueness, find the mean additive value function of those (near) optimal solutions which maximize the objective functions of (7), on the polyhedron of the constraints of the LP (6) bounded by the constraint of (8), where z^* is the optimal value of the LP in step 3 and ε a very small positive number.

$$u_i(g_i^*) = \sum_{j=1}^{a_i-1} w_{ij} \quad \forall i = 1,2,...,n \tag{7}$$

$$\sum_{k=1}^{m}[\sigma^{+}(a_k)+\sigma^{-}(a_k)] \leq z^* + \varepsilon \tag{8}$$

By applying the UTA* algorithm all the necessary parameters to estimate global utility functions $U(g(\alpha))$ for each alternative are calculated. Thus, a value is assessed quantifying alternative's utility to each user and ensuring consistency with his/ her value system.

Author's information

Kleanthi Lakiotaki is a PhD student in information and decision sciences at the Production & Management Engineering Department of Technical University of Crete. Her research interests include recommender systems, business intelligence, personalization technologies, decision support systems and multi-criteria decision analysis. She received her Bsc and Msc from University of Crete.

Contact her at the Dept. of Production and Management Engineering, Technical University of Crete, GR 73100, Chania, Crete, Greece; klio@ergasya.tuc.gr

Nikolaos Matsatsinis received his B.A. in Physics from Aristotle University of Thessaloniki (1980) and his PhD in Intelligence Decision Support Systems from Technical University of Crete (1995). He is a full Professor of Information and Intelligent Decision Support Systems and head of the Department of Production Engineering and Management of the Technical University of Crete, Greece. He is the author or co-author of eleven books and over of fifty eight articles in international scientific journals and books. His research interests fall into the areas of Decision Support Systems, Artificial Intelligent and Multi-Agent Systems, e-Business, e-Marketing, Multicriteria Decision Analysis, Group Decision Support Systems, Workflow Management Systems.

Contact him at the Dept. of Production and Management Engineering, Technical University of Crete, GR 73100, Chania, Crete, Greece; nikos@ergasya.tuc.gr

Alexis Tsoukiàs is a CNRS research director at LAMSADE, Université Paris Dauphine. He holds (1989) a PhD in Computer Science and Systems Engineering from Politecnico di Torino (Italy) where he also graduated engineering studies (1983). His research interests include subjects as: multiple criteria decision making, non conventional preference modelling, applied non classical logics, ordinal mathematical programming, ordinal measurement, multi-agent systems and software evaluation and selection. He has been teaching to several graduate and post-graduate classes in Politecnico di Torino and Université Paris Dauphine since 1987 and he was visiting professor at Technical University of Crete, Université Libre de Bruxelles and Rutgers University. He is a former President of EURO, the European Operational Research Association. Presently he is member of the CNRS National Council. He also coordinates the COST Action IC0602 Algorithmic Decision Theory.

Contact him at LAMSADE – CNRS, Universite Paris Dauphine, Place du Marechal de Lattre de Tassigny 75775 Paris France; tsoukias@lamsade.dauphine.fr