FRANCESCO RICCI AND FABIO DEL MISSIER

SUPPORTING TRAVEL DECISION MAKING THROUGH PERSONALIZED RECOMMENDATION

ABSTRACT

We present an approach to the design of personalized recommender systems that integrates content-based methods, collaborative filtering techniques and case-based reasoning while adopting a user-centered perspective. These techniques are employed to support information search and choice processes. In this framework, we developed and tested a system prototype (NutKing) that helps the user to construct a travel plan by recommending attractive travel products or by proposing complete itineraries. In the information search phase, the system aids the user in specifying a successful query that winnows out unwanted products in electronic catalogues and reduces the information overload. This is accomplished through two kinds of query rewriting operators (relaxation and tightening) in a mixed initiative approach. In the choice phase, the search results are sorted according to a case-base similarity metric, which takes into account the similarity between the users' travel preferences. The aim of this adaptive sorting is to highlight products that are potentially interesting, because they are similar to those selected by other users in an analogous context. The prototype has been empirically evaluated in a pilot study. The results of the pilot evaluation are discussed, with special reference to aspects concerning the user-system interaction aspects.

1.1 Travel decision making

Travel planning is a multi-faceted decision process consisting of choosing a destination and grouping together tourism products and services (attractions, accommodations, and activities) closely related to the destination (Dellaert, Ettema, & Lindh, 1998; Jeng & Fesenmaier, 2002; Moutinho, 1987).

The complexity of the concepts used and of the decision process involved in travel planning poses challenges for the design of usable and effective decision support tools. For instance, the terms *destination* and *travel plan* refer to fuzzy concepts that lack a commonly agreed definition. Furthermore, the spatial extension of a destination is known to be a function of the traveler's distance from the destination area. Italy could be a destination for Japanese, but a European traveler may focus more specifically on a particular region, such as Tuscany.

Moreover, a travel plan may vary greatly in its structure and contents, and different strategies can be used to construct it. For instance, some people may search

[Editor(s) here] (ed.), [Book Titlen here], 1—21.

for pre-packaged solutions (all-inclusive) while “free riders” may prefer to select each travel component separately. It is possible to start from a destination and then to search for other items (for example in a nature-oriented leisure holiday) or, alternatively, to focus on a particular activity or event (e.g., a conference or an exhibition) and then extend the plan taking into account the constraints brought about by this event. The same person may use different strategies in different contexts.

Travel decision making is one of the most comprehensively investigated areas in tourism research. In particular, many conceptual approaches to travel destination choice have been proposed. These approaches can be classified into four different frameworks: (1) choice set models (Crompton & Ankomah, 1993; Um & Crompton, 1990), (2) general travel models (Woodside & Lysonski, 1989), (3) decision net models (Fesenmaier & Jeng, 2000), and (4) multi-destination travel models (Lue, Crompton, & Fesenmaier, 1993). This literature classifies the variables used to predict the destination into the two broad categories of personal features and travel characteristics. Personal features include socioeconomic factors as well as psychological and cognitive traits. Travel characteristics comprise the situational variables that shape the travel, such as travel purpose, length the travel, distance to the destination, and travel group composition.

Despite the richness of travel decision making literature, only a very limited number of contributions have dealt with the topic of integrating decision models into travel recommender systems. This might be because the majority of existing models are based on traditional studies of consumer behavior, which are not focused on web technology or on travel interactive decision aids. These studies provide only general and limited guidance, because they do not take into account the unique characteristics and constraints associated with each specific communication medium and support tool. Therefore, trying to fill the gap between the travel decision models and “digitized” decision behavior is a valuable but difficult task, which requires the design and test of new models and aids.

1.2 Recommender systems

A recommender system helps the user to make choices when there is no sufficient personal experience of the available options. These kinds of systems can aid the consumer in various ways. They can simplify the information search process and facilitate the comparison of products (e.g., activebuyers.com), report the reviews of other users (epinions.com), or exploit the consumers’ history to suggest products similar to those purchased in the past or previously selected by users with a similar buying behavior (Amazon.com).

eCommerce web sites make use of recommender systems to suggest interesting and useful products and to provide consumers with information that is intended to support their decision processes (Kobsa, Koenemann, & Pohl, 2001; Schafer, Konstan, & Riedl, 2001). Recommender systems are mainly required in order to cope with information overload and lack of user knowledge in a specific domain. In general, they try to optimize some cost-benefit trade-off (for example, between the usefulness of the recommendation and the users' search and interaction costs).

Building real world recommenders demands a concerted effort and requires careful elicitation of user requirements, task analysis, development and tuning of the recommendation algorithms, and the design and testing of the graphical user interface.

Recommendation technologies are based on the implicit assumption that users' needs and preferences can be mapped into product selections, by employing the appropriate algorithms and the knowledge embedded in the system (Ricci, 2002).

Burke (2000) describes three different types of recommendation approaches: (a) collaborative-filtering or social-filtering, (b) content-based, and (c) knowledge-based. Here we will consider only the collaborative and the content-based methods, because they are integrated in our hybrid approach.

In the content-based systems, the user expresses some preferences on a set of products. Then the system retrieves from a catalogue the items that share common features with the products that have been judged interesting by the user. The results are typically sorted according to the degree of match with the user's preferences. The major drawback of the content-based approach is its excessive compliance with the preferences specified by the user, leaving no room for variability regarding the suggestions (O' Sullivan, Wilson, & Smyth, 2002).

In the collaborative-based approach (Breese, Heckermann, & Kadie, 1998), the system collects users' ratings on the suggested products and on the previously purchased items to infer any existing similarity between users. In this way the system can suggest to the user a set of novel products, which have been positively rated or previously bought by similar users. Collaborative approaches are effective in suggesting interesting items. Moreover, the quality of the recommendation improves over time, as new ratings are collected. On the other hand, collaborative-based methods require a huge number of ratings before producing satisfactory recommendations and they are not able to take into account specific (session-dependent) needs, since they uniquely rely on the users' past behavior and do not make any attempt to adapt the recommendations to the session-specific consumer's requirements. Collaborative filtering can be directly applied only to frequently purchased products (books, movies, or CDs). Other kinds of products (e.g., travel, cars) are far less amenable to a collaborative-based recommendation, especially as they are not purchased so often. Hence, the system cannot have an appropriate list of personal ratings and past ratings may not provide enough knowledge to predict the user's future choices.

1.3 Challenges for a travel recommender system

An effective and usable travel decision aid should flexibly support a range of different users' needs and various specific planning strategies. Conventional recommendation technologies have applied only general methods. Therefore, the main problem is to devise travel recommender systems that are able to satisfy the travelers' unique requirements by taking advantage of the knowledge and findings of travel decision theory. In our opinion, these systems should attempt to satisfy the following requirements:

– Products may have a complex structure, and the final recommended item should be a coherent aggregation of more elementary components. For instance a trip may be composed of a flight, an accommodation, and a ticket for a match.
– Both short-term preferences (related to a situation-dependent goal) and long-term stable preferences should influence the recommendation. Given that short-term preferences often arise from compelling needs, they should have greater weight than long-term preferences. For instance, if the user is currently searching for a business flight, the system must lessen the influence of a previous history of 'no frills' flights.
– The cognitive effort that the user devotes to the information search should be reduced but, in any case, the recommendations must satisfy the user's explicit preferences.
– The user and the system should collaborate in a mixed-initiative fashion, and each one must contribute what it does best. In particular, the user should be allowed to keep control of the interaction and to make informed choices, while the system must provide the relevant information (by offering easy ways to explore the option space and effortless means to specify and modify the information search).
– The lack of an initial database of user interactions or of a set of sales records should not prevent the application of the system support. Furthermore, both occasional and registered users should be able to obtain valuable recommendations.

Bearing in mind these requirements, we have conceived a methodology to support the information search and choice processes in the travel domain, based on the analysis of the travel decision making literature and of the existing recommendation technology. This methodology encompasses both the interaction design (i.e. how the interaction is structured and sequenced) and the different kinds of support provided in the different interaction stages (i.e. the specific decision aid provided for each interaction step). The interaction design cannot be considered as neutral, but it is always a means of shaping the decision process according to our hypotheses about the appropriate ways to aid the user. In our case, we have tried to support a two-stage decision process, with an initial noncompensatory screening of the options and a subsequent choice phase focussed on a set of attractive alternatives. This process may sometimes result in sub-optimal choices (Edwards & Fasolo, 2001), but it is coherent with the user's choice strategies when the initial number of options is high (Payne, Bettman, & Johnson, 1993). Moreover, it allows for a reduction in the cognitive effort associated with the specification of preferences and with the information search.

In section 2 we will present the interaction design and in section 3 the technologies that we have developed to enable the proposed interaction. Section 4 illustrates the pilot evaluation of the NutKing prototype. Finally, in section 5, we will discuss issues related to the topic of personalized decision support in the travel domain and sketch some directions of our future work.

2. INTERACTION DESIGN

In this section, we will first describe the interaction design in abstract terms. Then we will provide a concrete example in order to make the stages that the user goes through more explicit. The example describes a real interaction session with a system prototype (NutKing), which is a specific implementation of our support methodology. In the next section we will present the techniques used to provide specific aids in the different interaction steps and we will summarize their theoretical underpinnings.

A recommendation session is divided into three stages:

– *Stage 1: Acquisition of travel preferences.* The system asks the user to provide some information about personal and travel characteristics. These features include group composition, means of transport, type of accommodation, budget, travel period, knowledge of the destination area, and preferred activities. These characteristics are considered as predictors of the user's decision behavior by the models of destination choice. They are called *collaborative features* and used by the system to retrieve similar *recommendation sessions* from a case base. The retrieved cases will be used, in the subsequent stages, to generate recommendations and to sort the results of the user's searches.
– *Stage 2: Search for travel products.* The user starts the process by seeking a destination or a product (accommodation, event, sport activity) that satisfies a set of requirements (for example, a destination close to a lake with golf facilities, or a budget hotel near the city center). The *content features* are used to define these requirements. When the search is successful, the system presents a set of results that satisfy the requirements. When the number of results is high, the system asks the user to provide some additional content features in order to winnow out the less interesting products (*tightening* of the query). In the case of a search failure (no results), the system asks the user to withdraw some content features, suggesting a set of minimal query modifications that will produce some results (*relaxation* of the query).
– *Stage 3: Choice.* In the last stage, the results obtained from a successful query are sorted and presented to the user. The recommender system finds similar recommendation sessions, employing all the collaborative features previously entered by the user. After the retrieval of a small set of similar sessions, the system sorts the list of results according to a specific similarity ranking criterion, which takes into account the agreement between the users' travel preferences (see section 3). The aim of this operation is to highlight products that are potentially interesting, because they are similar to those selected by other users in similar recommendation sessions. Once the items have been presented to the user, their features can then be carefully examined and the user may decide to add some products to a shopping cart (called *travel plan*). Obviously it is

possible (and quite frequent) that the user cycles through the second stage (search) and the third stage (choice), modifying the search criteria and examining new sets of results.

Our approach enables content features to dominate the selection process and to determine which items will compose the choice set. But the collaborative features are exploited to rank the set, highlighting the most attractive items.

A quite different way to use the system is to exploit the so-called *complete travel recommendation* function, which produces results in the form of whole travel packages (see section 3). In our prototype, this option is currently available as an alternative to the search for travel products (stage 2).

We will now present an example to illustrate our abstract description. Imagine you are a European tourist, mainly interested in culture and nature. You are planning your first trip to Trentino (a region of northern Italy), and decide to use the NutKing system (http://itr.itc.it) to find some advice regarding this trip. After a welcome page, the recommender asks you to provide some information about your travel preferences (Figure 1). You are required to fill out a form, and the collaborative features are thus acquired (stage 1).

Figure 1: Acquisition of travel preferences.

During stage 2 you start searching for travel items. After deciding that you will visit Trento, you seek a particular type of accommodation in the Trento area (for example a hotel between 2 and 4 stars, within a given price range: figure 2 on the left). The search produces 16 results and the system suggests that some constraints could be added (tightening), while also providing a list of content features that may be specified (figure 2 on the right).

Figure 2: Search for an accommodation and tightening.

You decide to add the following constraint: the hotel should be near the city center. The search is performed again and it produces eight results, which are sorted according to the system ranking criteria (stage 3: figure 3). After a careful examination of the information available on the various options, you add a hotel to your travel plan.

Figure 3: List of results.

Now you begin to search for a destination in the southern areas. You would like it to offer many different things, such as historical sites, museums, folklore, jazz music, horse riding, and cycle paths. You specify a query (figure 4 on the left), but you get no results: it seems that you were asking too much! The system suggests you remove some constraints (relaxation: figure 4 on the right). If you avoid specifying the area you will get two results, while removing 'horse riding' or 'cycle paths' will produce a single result. You decide to withdraw horse riding and then the city of Rovereto shows up.

Figure 4: Search for a destination and relaxation.

Figure 5 depicts the structure of the interaction in stages 2 and 3.

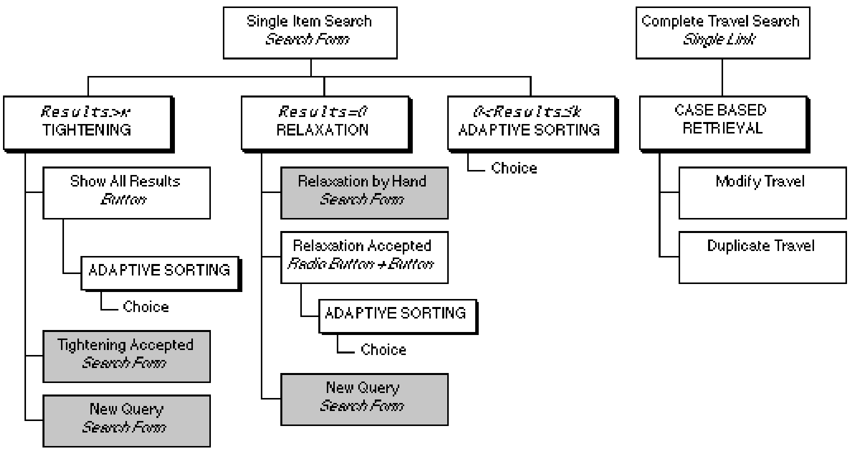

Figure 5: The possible NutKing interaction paths (stages 2 and 3). The shadowed boxes mark the system intervention. The gray boxes indicate the composition of a new form-based query or the modification of one previously formulated.

In stage 2, the user has two main options. The first one is to request a complete travel recommendation by following a link. Alternatively, it is possible to compose a query to seek a specific type of travel item (a location, an accommodation, an event, sports). This single item search is performed by supplying content features in specific search forms.

If a complete package is requested, the suggested travel items can be modified or duplicated, in order to transfer some information to the user's personal travel plan.

When the user searches for a single item, the search can produce three possible outcomes: (a) no results, (b) more than k results (k is a threshold with a default value of 10), and (c) a number of results between 1 and k (inclusive).

If the recommender realizes that the query will not produce any results, it suggests to the user to relax some constraints. This can be done simply by selecting the preferred suggestion or by refining the query by hand on the single item search form. Alternatively, the user can enter a new query (i.e., a query that cannot be considered as a relaxation of the previous one).

If the query produces many results, the recommender proposes to the user some ways to tighten the search. After the tightening suggestion, the user can choose from three options: (a) follow the system advice and modify the query accordingly, (b) display all the results, or (c) formulate a new query (i.e., a query that cannot be considered as a tightening of the previous one).

The search results are always sorted according to the specific similarity measure employed by the system, and they are subsequently presented to the user (three items per page).

Finally, the user examines the result pages, acquiring information about the options, and an item can be selected to be added into the travel plan.

A typical interaction session is composed of several queries, and each unsuccessful query can be refined in an iterative way.

3. RECOMMENDATION METHODOLOGY

Trip@dvice, the proposed recommendation methodology, satisfies the requirements summarized in subsection 1.3, while also supporting the interaction process described above. It integrates case-based reasoning (CBR), cooperative query answering, and collaborative filtering. The logical architecture of the recommendation methodology is shown in Figure 6. The recommendation manager is the entry point for all of the recommendation functions and is accessed via the graphical user interface. The other modules will be described in the following sections, starting with the case-based reasoner.

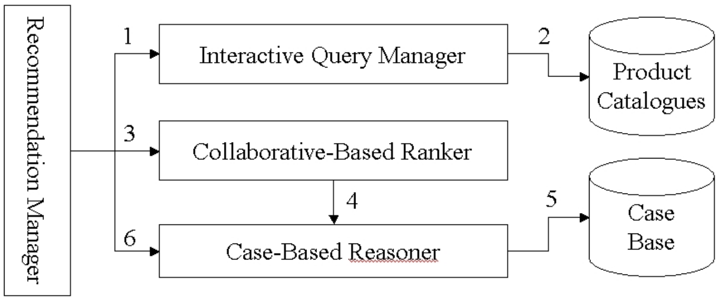

Figure 6: Logical architecture of the recommender system.

3.1 Recommendation sessions as cases

CBR is a multi-disciplinary research area, focused on the exploitation of past experience encoded in the form of cases (Aamodt & Plaza, 1994; Aha, 1998). There are at least two different conceptions of CBR: it is regarded as a plausible high-level model for cognitive processing (Kolodner, 1993) or as a computational paradigm for problem solving (Aamodt & Plaza, 1994). We adopt the latter view, referring to CBR as a problem solving paradigm that relies on the specific knowledge gathered by solving practical problems. In CBR, a case comprises two main components: the problem and the description of the solution. The basic assumption of CBR is that similar problems share similar solutions. In our approach, a case models a unique user-system recommendation session and it comprises:

- the travel preferences provided by the user in the first stage of a recommendation session (i.e., the collaborative features);
- the products chosen by the user and belonging to the travel plan;
- the user's profile, composed of some demographic and personal data (age, gender, nationality, address, etc.); this is available only if the user is registered.

In our model, the collaborative features provided by the user (and the user profile) describe the "travel problem" that the user "solved", whereas the chosen products represent the solution. A set of recommendation sessions involving many different users is stored in the case base, while the catalogues of tourism products and services (attractions, accommodations, activities) are stored in several databases.

A straightforward application of classical CBR solves a recommendation problem by searching for similar recommendation sessions. This means that the problem definition (collaborative features) is used to retrieve cases (matching), and that the system suggests to the user products which are contained in the retrieved cases. In our system, this simple approach is adopted in the *complete travel recommendation* function. This function is implemented by asking the CBR manager to create a case, initialized with the user-provided collaborative features. The newly generated case represents the current recommendation session (arrow 6 in figure 6) and it is used to retrieve a set of similar cases from the case base (arrow 5 in figure 6).

This approach has limitations. First, the products found in the retrieved cases may be temporarily unavailable: the suggested hotel may be fully booked and a given event may have already taken place. Second, this approach does not consider specific user preferences (for example, the user may be interested in a budget hotel close to the city centre). Searching for cases with additional constraints (containing only the available products with the desired features) would almost certainly produce a failure, unless the case base contained a wide range of offers. This ideal condition seems very difficult to attain, because it requires devoting a long time for the acquisition of cases (bootstrap) before the system can deliver useful recommendations.

Therefore, we have decided to use in a different way the implicit preferences that derive from similar cases. In our approach, these past recommendation sessions are utilized to rank the results of the searches that the user has performed in the product catalogues (the details of the ranking algorithm are described in section 3.3). In the current implementation, the system dynamically builds a user model for each recommendation session. This model comprises a set of cases retrieved from the case base. These cases own vectors of collaborative features similar to the vector representing the current case. The similarity between the current case and a retrieved case is defined by a heterogeneous distance metric that is computed as a weighted sum of local distances (i.e. feature differences, Ricci et al., 2003). At present, the user profile is not taken into account in the generation of the user model but we are planning to expand the approach to include these personal features.

3.2 Interactive query management

The goal of the cooperative query answering research area (Gaasterland, Godfrey, & Minker, 1992; Godfrey, 1997) is to create systems that attempt to understand the meaning of a user query, suggest or answer related questions, and produce an answer from the available data or give an approximate response.

In our approach, we apply some methods borrowed from this area when a user is searching for products that satisfy some preferences expressed using descriptors of the product (the content features specified in the second stage of the process previously described). For instance, if the user is looking for a budget hotel close to the center of the city of Trento, the Interactive Query Manager Module (IQM) performs the following steps before displaying any recommendation:

1. IQM searches for all those hotels that satisfy the user-specified constraints (arrow 2 in figure 6).
2. IQM analyses the result set.
   a) If no result is found, the system computes an alternative version of the user query for each specified content feature, relaxing the constraints contained in the initial user query. Hence, in the example mentioned before, two queries are generated: the first relaxes the budget condition while the second relaxes the preference on the hotel location. Then the system offers the user a choice among the relaxed queries that result in a non-void set of items. This choice is

presented with the explanation that it is not possible to satisfy all the constraints specified in the initial query, but that there is the possibility to find some products that partially satisfy the original request (Ricci, Mirzadeh, & Venturini, 2002). The aim of the relaxation function is to maximize the probability of finding some results, taking into account both the user's preferences and the composition of the product catalogue.

b) When a query produces too many items (above a threshold value) it is assumed that the user will not browse the entire result set to locate the most interesting items. Therefore, we have designed a system function that explains this situation of potential information overload and lets the user choose between displaying all the results and tightening the original query. In this last case, the system suggests some tightening options by identifying new product features that are potentially interesting but not yet specified. For instance, if the user searched for a hotel in Trento, the system asks the user to specify the desired values for an additional content feature, as there are many hotels in Trento. It does this by displaying a list of alternatives (for example, the price range, the category, or the distance from the city center). The list of the features that can be used for tightening a specific query is established by an unsupervised feature selection method that combines the information gain of a feature and its popularity (Ricci et al., 2003). Those features most frequently used in queries of the same type as the one currently specified are selected as candidates, if they have not been already specified. Otherwise, the system selects the unused feature with the highest entropy value. Entropy is a measure of "disorder" of the feature values: the higher the entropy the more uniform is the distribution of values. Hence, assigning a value to the feature with the highest entropy will make it more likely that the new result set will be smaller. For instance, if only half of the hotels are close to the city center but almost all of the hotel rooms are equipped with a TV, then the user will be asked to specify a value for the "close to center" feature.

The proposed interactive query management function enables the user to perform a more efficient search from the available options, implementing a mixed-initiative scheme. The system is better suited to compute the alternative successful relaxations or the best candidate features for query tightening, while the user can choose which query change is more satisfying.

3.3 Collaborative-based sorting

A custom implementation of the collaborative filtering approach is applied in order to present the result list of products obtained from a user query. This function is used at the beginning of the third interaction stage (choice) after the potential intervention of the cooperative query answering methods (in the second stage).

In classical automated collaborative filtering the system rates the products which have not yet been judged by the user by taking a weighted average of the ratings of similar users (Breese, Heckerman, & Kadie 1998; Schafer, Konstan, & Riedl, 2001).

The weights in this approach are the similarity values, where the similarity between two users is computed by taking the correlation of their ratings.

In our approach, we want to support even users who have no rating history or who are not interested in logging into the system. Moreover, the travel decision choice literature points out that the specific features of the current travel plan should be appropriately taken into account. For the above reasons, the preferences associated with the specific travel products that the user is currently searching should play a prominent role. On the other hand, as the success of the recommender technology demonstrates, it would seem useful to exploit also the implicit knowledge that derives from the information about similar users. As previously described, we have combined the two approaches: after the user has obtained a reduced set of candidates through a content-based search, the collaborative-based ranking method is applied. The reduced set of options loosely corresponds to the *late consideration set* of the choice set models (Crompton & Ankomah 1993; Um & Crompton 1990): the decision maker will acquire information about a subset of these options that, given the current constraints, are deemed attractive (the *action set*).

The collaborative-based ranking is computed in the following way. A small set of recommendation sessions (cases) is retrieved using a similarity function that takes into account only the collaborative features of the current case (arrow 4 in figure 6). These sessions have similar values for the predictive features and therefore they should provide good recommendations for the current case. The rationale is that since the current session and a retrieved session both share the values of the predictive features (problem definition in CBR terminology) they should also share the recommended solution (i.e. the destination or the tourism products contained in the travel plan). The items of the reduced set of candidates that are more similar to the items contained in the retrieved sessions are considered as good recommendations. These items are presented to the user in a sorted list. The order of the list reflects the estimated values of the recommendations for the user, and these recommendation values are displayed explicitly on a graphical five-point scale.

The reduced set of candidates is actually sorted according to a score function (Ricci et al., 2003) that is computed by multiplying the session similarity (between the current case and one that has been stored) by the item similarity (between an item in the reduced set and an item in a similar case). For both the similarity functions we have used the Heterogeneous Euclidean Overlap Metric (HEOM) as a basis (Wilson & Martinez, 1997). The distance between two n-dimensional vectors including both numeric and symbolic features, $d(v,u)=d((u_1, \ldots, u_n), (v_1, \ldots v_n))$, is computed as the square root of the sum of squares of the local feature distances $d_i(u_i,v_i)$. The local distance is given by the module of the difference of values for numeric features and by the equality test, $d_i(u_i,v_i)=0$ *if* $u_i,=v_i$, for symbolic features. HEOM distance is normalized and ranges between 0 and 1, whereas similarity is computed as $1\text{-}d(v,u)$.

4. PILOT EVALUATION

We will now describe the first pilot evaluation of the system prototype, outlining the main results and empirical findings, and focussing our attention on a detailed analysis of user-system interaction.

This study provided three main contributions:

- it gave some preliminary indications regarding the specific strengths and weaknesses of the system, and on the interaction design;
- it suggested some hypotheses that were able to promote our understanding of the user-system interaction;
- it provided significant methodological feedback, which will help us to plan further evaluation studies.

4.1 Pre-evaluation qualitative assessment

In accordance with the viewpoint of user-centered design, we took into account the users' needs from the first stages of development. System requirements were derived from the analysis of the travel decision making literature and from the identification of usage scenarios.

A series of qualitative evaluations (cognitive walkthroughs, heuristic evaluations, and detailed analyses of the interaction structure) have been carried out in various development stages by different experts, promoting significant changes to the design of the system and of the user interface.

4.2 Evaluation goals and system limitations

The NutKing prototype has been evaluated in order to acquire some preliminary feedback on the system efficiency and effectiveness, and to obtain some indications on the quality of the interaction design. We were also interested in obtaining some evidence on the statistical power of the tests and on the evaluation procedure, which is useful for blueprinting further experimental studies.

It should be noted that the NutKing user interface was a prototypical version that had never undergone an empirical test. Furthermore, the case base of the system was very small (35 cases). These cases had been generated by a group of expert and naive users.

The NutKing product catalogues were comprised of locations (220 records and 26 attributes), accommodations (1618 records and 20 attributes), sport activities (1624 records and 7 attributes), events (3286 records and 5 attributes), and cultural attractions (538 and 3 attributes). These data types are views over an Oracle database listing more than 100 tables, developed by the regional destination management organization (APT Trentino).

4.3 Experimental design and procedure

Using a between subjects design, the participants were randomly assigned either to the NutKing+ or to the NutKing- group. The first group (n=16) used the full functionality recommender (NutKing+), while the second group (n=19) interacted with a baseline system (NutKing-), deprived of the interactive query management and of the adaptive sorting functions. The two system variants (and interfaces) were very similar and the interaction flow was basically the same: a cycle of preference acquisition, query generation, and product choice.

The participants, students and office workers of the University of Trento, had to plan a vacation in Trentino, searching for appropriate travel items and selecting a set of products. Given the exploratory nature of the study, there were no specific constraints on the accomplishment of the task and all of the participants were familiar with the target area. The participants were told only that the experiment was about "travel planning" and that they could work at their own pace. They were not given any hint as to the type of system that they would use, and we tried to avoid generating any kind of expectation or Hawthorne effects (Shadish, Cook & Campbell, 2002).

The experiment was preceded by a training period (about 10 min), during which the participants were allowed to freely explore the system and to read the on-line explanations. After the experimental session, the participants completed a tailor-made evaluation questionnaire, and they were asked to express their comments and observations on the system.

The two main hypotheses were as follows:

- (H1) the NutKing+ system will be able to provide useful recommendations (the items selected by the NutKing+ participants will be placed closer to the top of the result list than the items picked up by the NutKing- participants);
- (H2) the use of the NutKing+ recommender produces an improvement in search and decision efficiency (provided that there are a similar number of items collected in the travel plans, the number of queries and the total amount of time taken will be lower in the NutKing+ system).

4.4 General results

The results of the experiment are summarized in table 1.

The participants of both groups specified a similar amount of input information (on identical forms). Both the mean number of collaborative features (NutKing+: *M*=11.5, *SD*=2.1; NutKing-: *M*=12.3, *SD*=1.4) and the mean number of query conditions (NutKing+: *M*=4.4, *SD*=1.1; NutKing-: *M*=4.7, *SD*=1.2) are very similar. t-tests with separate variance estimation and Mann-Whithney U tests did not show statistically significant differences. Also taking into consideration that no clear differences were found in the type of information specified, it appears that the system type did not affect the query generation process.

Table 1: Mean scores for the two NutKing variants. The variables marked with * are associated with statistically significant differences (p<.05).

Variable	NutKing+	NutKing-
Number of collaborative features	11.5	12.3
Number of query conditions	4.4	4.7
Number of queries	13.4	20.1
Number of results *	9.8	42
Number of pages displayed	71.3	93.3
Session duration (min)	31	28.5
Number of items in the travel plan	4.1	5.8
Mean position of the selected items in the result list *	2.2	3.2

NutKing- users formulated more queries than NutKing+ users, but this difference is not statistically significant. The Levene test showed that the variance in the number of queries is significantly lower for the NutKing+ system (F(1,33)=4.29, p<.01; NutKing+: SD=9.25; NutKing-: SD=19.17). The mean number of results produced by NutKing- is greater than the mean number of results produced by NutKing+ (t(33)=2.05, p<.05). A similar trend is observed in the mean number of pages displayed by the two systems, and the variance of the number of results is again significantly lower in NutKing+ than in NutKing- (Levene test: F(1,33)=18.39, p<.0001).

Despite some differences in the queries and in the results, the session duration is not significantly different in the two conditions. Therefore, it seems that NutKing+ users have spent more time examining information rather than composing queries or moving through the pages. Thus we obtained some partial indications that NutKing+ can promote efficiency (H2), but it can be hypothesized that the time saved is devoted to a deeper analysis of the information and to the choice processes.

The travel plans assembled using the two systems are composed of a similar number of items, but the mean position of these items in the result list is significantly closer to the top in NutKing+ (t(141)=1.96 p=.05 with log transformed data). In general, the participants tend to select the upper items in the result list. However, in accordance with H1, NutKing+ participants picked those items with higher recommendation rankings.

4.5 User-system interaction

We analyzed the NutKing+ log data, our main goal being to try to understand if the interaction design was appropriate and to acquire some indications relating to the participants' interaction strategies. We took into account 216 queries (all those queries with complete information) and summarized the observed interaction paths (figure 7).

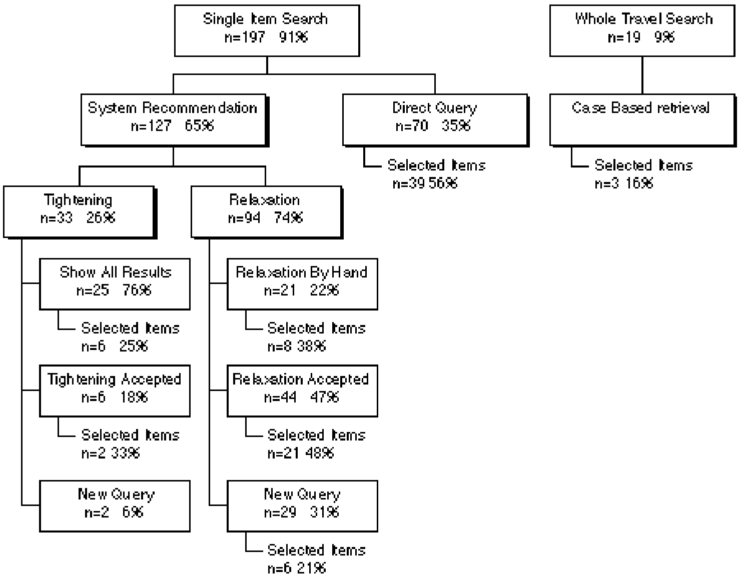

Figure 7: The observed interaction with NutKing+. The frequency of each action and its occurrence percentage, computed in relation to the alternative actions, are presented inside the action box. The frequency of item selection (i.e. adding an item to the travel plan) and the item selection percentage after each specific action are shown below the corresponding action box.

The complete travel recommendation function was rarely used (19 queries, 9%). Furthermore, the travel product suggested by the system was accepted by the participants in only 3 cases, and the difference between the proportions of selected items after a single item search and after a complete travel request is statistically significant (single item=.42, complete travel=.16, $p<.05$). We can formulate many hypotheses to explain these results. The more reasonable are as follows: (a) an inadequate interaction or interface design (i.e. the complete travel recommendation was not apparent in the interface or its meaning was not made clear), (b) the unsatisfying nature of the specific system suggestions with regard to the whole travel search, and (c) the participants' general unwillingness to rely on the suggestion of a complete travel package. Further experiments should contrast these hypotheses, given that the participants' free observations did not provide specific indications.

The single item search was quite frequent (197 queries; confidence interval for the proportion: .87=.91=.94), and it was repeated in an iterative way until the travel plan was deemed complete.

The system proposed a query change in 65% of single item queries. In the majority of cases, the initial request was too specific (74%), and the users accepted the system's relaxation hint. This subsequently led to the selection of a travel item in

31% of cases. So the participants accepted the system suggestion, and this proved to be useful: the proposed relaxations produced a significantly higher proportion of item selections than a different type of query change (relaxation=.45; other query change=.21, $p<.05$). It should also be noted that the relaxation interaction has probably not being designed perfectly, because some participants have relaxed their queries 'by hand' (cf., section 2).

The tightening function was rarely used: the participants preferred being able to examine all the results (95% confidence interval for the proportion: 0.59=0.76=0.88). Again, many hypotheses can be put forward: (a) the participants do not want to winnow out good options without having the possibility to actually view them, (b) the participants like to explore the option space, and (c) the participants avoid the cognitive effort associated with the tightening operations. As in the previous case, only additional research would help to provide the answer.

4.6 Discussion

The pilot study provided some weak indications that the NutKing+ variant appears to be able to provide useful recommendations. The use of the recommendation-enhanced system also changed the users' information-seeking patterns and probably produced a deeper analysis of the information retrieved from the search. In any case, the real effectiveness of the decision support provided by the system must be further investigated through more powerful tests, encompassing a broader and more detailed set of measures (for example: more detailed log data, satisfaction ratings on the products, high-quality post-study questionnaires). The execution of web experiments (Birnbaum, 2000) as a complement to the traditional laboratory study could help us to increase the external validity of the evaluation, even if this approach will require solving many methodological and technical issues.

Specific indications have been acquired on the support functions. The relaxation support seems to be useful, but it requires a better interface design. The tightening support and the complete travel recommendations should probably undergo a deeper reconsideration, once a complete understanding of the reasons underlying their low usage rate in our specific decision environment is established.

5. CONCLUSIONS AND FUTURE WORK

Our experience shows that it is possible to design personalized applications for the travel domain. It is quite a difficult task, but the potential benefits could be very high, given the huge variability of users' preferences in this domain.

We have built a prototype system that is able to satisfy a series of basic requirements (including some partial form of personalization), and we are currently improving its design and exploring alternative recommendation possibilities. We decided to follow this 'requirement-oriented' approach (instead of adopting a model-based approach) because of the existence of many problems that must still be solved in order to be able to fully specify other kinds of decision aids for the tourism domain. The first problem is the lack of detailed knowledge about the users' interactive information search processes and decision behaviors. We do not have sufficient knowledge of the tourists' decision processes to be able to fully specify satisfying descriptive models. This knowledge should concern a detailed analysis of the ways in which different types of tourists use the different media and decision aids in their travel planning efforts. A second problem is related to the difficulty of translating some existing general-purpose prescriptive approaches into real decision aids. In order to do this, we will need to apply a trusted normative model, such as the multiattribute utility theory (Keeney & Raiffa, 1976). These normative models are not easily applicable to the travel domain and their 'embodiement' in this domain will require the solution of complex problems (for example, the selection of the attributes for shaping the travel decision problem). Therefore, we have devoted most of our efforts to the reduction of the search and interaction costs and we have attempted to exploit the implicit knowledge stored in memorized recommendation sessions.

Some significant lessons learned from the design of the NutKing system and from the review of the first pilot evaluation are as follows:

- The design of a travel recommender system requires taking into account preferences regarding different types of products. Much attention should be given to the interaction design in order to keep the multiple-item search process as simple as possible and to allow a seamless integration of the explicit preferences and of the collaborative recommendations.
- The real usefulness of the main support functions should be empirically tested. In our pilot study we observed that some functions were rarely used (tightening and complete travel recommendation). Our current evidence is solely observational, but experimental tests could be performed to analyze the usefulness of specific system functions.

Our future work will follow three main directions:

- We will improve in various ways the recommendation technology described in this paper, increasing its degree of personalization. For instance, we will enable the learning (adaptation) of the feature weights used in the similarity measures, employing a punishment/reward algorithm (Ricci & Avesani, 1999). Moreover, we will try to offer a seamless support to users who adopt different decision styles (Fesenmaier, Ricci, Schaumlechner, Wöber & Zanella, 2003).
- The potential of different methods will also be explored. In particular, we are interested in some approaches deriving from the behavioral decision making literature (Edwards & Fasolo, 2001) that prevent the users from adopting sub-optimal noncompensatory decision strategies. As we have just pointed out, applying these approaches to the tourism domain will require specific research effort and carefully designed support tools.
- Finally, we are currently trying to define an integrated approach for the evaluation of recommender systems and decision-facilitating web sites for consumer decision making (Del Missier & Ricci, 2003). This approach will be based on a principled integration of focussed experimental tests of specific interface and system components, simulations, laboratory experiments within a

context-matching approach (Payne, Bettman, & Schkade, 1999), and web experiments (Birnbaum, 2000). The insights acquired from the layered evaluation method will be taken into account in the design of the experiments (Karagiannidis & Sampson, 2000). A preliminary step will be the identification of the appropriate evaluation dimensions for the different types of decision aids (Yates, Veinott, & Patalano, 2003).

6. ACKNOWLEDGEMENTS

The research was supported by CARITRO Foundation, European Union (contract DieToRecs IST-2000-29474), and a grant from Provincia Autonoma di Trento (Indirecs project).

REFERENCES

Aamodt, A., & Plaza, E. (1994). Case-based reasoning: foundational issues, methodological variations, and system approaches. *AI Communications*, 7, 39-59.

Aha, D. W. (1998). The omnipresence of case-based reasoning in science and application. *Knowledge-Based Systems*, *11*, 261-273.

Birnbaum, M. H. (2000). *Psychological experiments on the Internet.* San Diego: Academic Press.

Breese, J., Heckerman, D., & Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering. *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, 43-52.

Burke, R. (2000). Knowledge-based recommender systems. In J. E. Daily, A. Kent, & H. Lancour (Eds.), Encyclopedia of Library and Information Science, Vol. 69, Supplement 32. New York: Marcel Dekker.

Crompton, J. L., & P. K. Ankomah. (1993). Choice set propositions in destination decisions. *Annals of Tourism Research*, *20*, 461-476.

Del Missier, F., & Ricci, F. (2003). Understanding recommender systems: Experimental evaluation challenges. *Proceedings of the Second Workshop on Empirical Evaluation of Adaptive Systems*, 31-40.

Dellaert, B. GC., Ettema, D. F., & Lindh, C. (1998). Multi-faceted tourist travel decisions: a constraint-based conceptual framework to describe tourists' sequential choices of travel components. *Tourism Management*, *19*, 313-320.

Edwards, W., & Fasolo, B. (2001). Decision Technology. *Annual Review of Psychology*, *52*, 581-606.

Fesenmaier, D. R., & Jeng, J. (2000). Assessing structure in the pleasure trip planning process. *Tourism Analysis*, *5*, 13-17.

Fesenmaier, D. R., Ricci, F., Schaumlechner, E., Wöber, K., & Zanella, C. (2003). DIETORECS: Travel advisory for multiple decision styles. In A. J. Frew, M. Hitz, & P. O'Connors (Eds.). *Information and Communication Technologies in Tourism 2003* (pp. 232-241). Springer-Verlag.

Gaasterland, T., Godfrey, P., & Minker, J. (1992). An overview of cooperative answering. *Journal of Intelligent Information Systems*, *1*, 123-157.

Godfrey, P. (1997). Minimization in cooperative response to failing database queries. *International Journal of Cooperative Information Systems*, *6*, 95-159.

Jeng, J. & Fesenmaier, D. R. (2002). Conceptualizing the travel decision-making hierarchy: a review of recent developments. *Tourism Analysis*, 7, 15-32.

Karagiannidis, C., & Sampson, D. (2000). Layered evaluation of adaptive applications and services. In P. Brusilovsky, O. Stock, & C. Strapparava (Eds.), *Adaptive hypermedia and adaptive web-based systems*, Lecture Notes in Computer Science, Vol. 1892, (pp. 343-346). Springer-Verlag.

Keeney, R. L. & Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value trade-offs*. New York: Wiley.

Kobsa, A., Koenemann, J., & Pohl, W. (2001). Personalised hypermedia presentation techniques for improving online customer relationships. *The Knowledge Engineering Review*, *16*, 111-155.

Kolodner, J. (1993). *Case-Based Reasoning*. San Mateo: Morgan Kaufmann Publishers

Lue, C. C., Crompton, J. L., & Fesenmaier, D. R. (1993). Conceptualization of multi-destination pleasure trip decisions. *Annals of Tourism Research*, *20*, 289-301.

Moutinho, L. (1987). Consumer behavior in tourism. *European Journal of Marketing*, *21*, 2-44.

O' Sullivan, D., Wilson, D., & Smyth, B. (2002). Improving case-based recommendation, a collaborative filtering approach. *Proceedings of the Seventh European Conference on Case-Based Reasoning*, 278-291.

Payne, J. W., Bettman J. R., & Johnson, E. J. (1993). *The adaptive decision maker.* New York: Cambridge University Press.

Payne, J. W., Bettman, J. R., & Schkade, D. A. (1999). Measuring constructed preferences: Towards a Building Code. *Journal of Risk and Uncertainty*, *19*, 243-270.

Ricci, F. (2002). Travel recommender systems. *IEEE Intelligent Systems*, *17*, 55-57.

Ricci, F., & Avesani, P. (1999). Data compression and local metrics for nearest neighbor classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21, 380-384.

Ricci, F., Mirzadeh, N., & Venturini, A. (2002). Intelligent query managment in a mediator architecture. In T. Samad & V. Sgurev (Eds.), *First International IEEE Symposium on Intelligent Systems* (pp. 221-226). IEEE Press.

Ricci, F., Venturini, A., Cavada, D., Mirzadeh, N., Blaas, D., & Nones, M. (2003). Product recommendation with interactive query management and twofold similarity. *Proceedings of the Fifth International Conference on Case-Based Reasoning*, 479-493.

Schafer, J. B., Konstan, J. A., & Riedl, J. (2001). E-commerce recommendation applications. *Data Mining and Knowledge Discovery*, *5*, 115-153.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs*. Boston: Houghton Mifflin.

Um, S., & Crompton, J. L. (1990). Attitude determinants in tourism destination choice. *Annals of Tourism Research*, *17*, 432-448.

Yates, J. F., Veinott, E. S., & Patalano, A. L. (2003). Hard decisions, bad decisions. On decision quality and decision aiding. In S.L. Schneider & J. Shanteau (Eds.). *Emerging perspectives on judgment and decision research* (pp. 13-63). New York: Cambridge University Press.

Wilson, D. & Martinez, T. (1997). Improved heterogeneous distance functions. *Journal of Artificial Intelligence Research*, *11*, 1-34.

Woodside, A. G., & Lysonski, S. (1989). A general model of traveler destination choice. *Journal of Travel Research*, *27*, 8-14.