

# Intelligent Hybrid Architecture for Tourism Services

Ana Almeida<sup>1,2</sup>, Bruno Coelho<sup>1</sup>, Constantino Martins<sup>1,2</sup>

<sup>1</sup>GECAD - Knowledge Engineering and Decision Support Group

<sup>2</sup>Computer Science Department

Institute of Engineering – Polytechnic of Porto

R. Dr. António Bernardino de Almeida, 431. 4200-072 Porto, Portugal

{amn, besc, acm}@isep.ipp.pt

**Abstract.** This paper presents an artificial intelligence-based architecture to be used in various tourism services, namely the user modeling and recommender system components. Through the use of several machine learning techniques, such as linear models, neural networks, classification and even text mining, a hybrid and complete approach at understanding tourism application users is attained. To assess the advantages of the model, a specific prototype was created, including a complex user modeling process and an advanced and innovative recommender system. The system was tested in the scope of Oporto city, in Portugal.

**Keywords:** Recommender systems; tourism; user modeling

## 1 Introduction

The tourism domain, and in particular, the holiday choosing process, represent very complex decision-making matters. The user is generally faced with the endless group of existing options, apart from the heterogeneity of the places to visit, for example, the type of Points of Interest (POI) available, hosted events and so on.

Such a scenario is clearly electable for the use of Artificial Intelligence (AI), and in particular, Recommender Systems (RS) [11], in order to filter the item search space. Such systems require the use of a coherent user model in order for results, and the overall system, to be customized and targeted to him / her: that's where User Modeling (UM) techniques come into play [1]. After an assessment of the current tourism systems, as well as other systems which use RS, there is the belief that significant work can still be done, regarding a more complex modeling of users, the introduction of innovative filtering techniques and ultimately more intelligent domain item recommendations.

This paper is organized as follows: section 2 will present a brief approach regarding RS in tourism, while section 3 presents the RS itself. Within this section, all different techniques that come into play when the RS is fired are thoroughly depicted. In section 4, the developed tourism application will be presented regarding its advantages and testing results. Finally, in section 5, some conclusions will be outlined, along with future work perspectives.

## 2 Recommender Systems in Tourism

Despite recent evolutions in RS, current systems still do not explore the huge potential of this area, namely in the tourism domain. User information, needed for the filtering effort, is extensively requested at startup, without necessarily being used throughout the remaining life-cycle of the application. Most of the times, though, systems rely on single and / or poor representations of the assumptions they make about users, which ends up in incoherent recommendations being performed [3][15]. A few examples follow:

- FilmTrust [14]: FilmTrust is a web-based system that explores the concept of trust in a movie related social network. In FilmTrust, users can not only express their particular opinion about a movie, but also define a trust degree for other users and their opinions. This follows the principle of basing predictions on reliable peers, instead of solely on similar ones.
- TripAdvisor [18]: this tourism website advises locations and activities for users, also containing a highly social component. Although this application contains one of the most hyped Recommender Systems (RS). UM does not seem to make a great part of the system's philosophy, being replaced by purely social theories;
- WAYN [19]: this application is an evident proof of a Web 2.0 social endeavor regarding tourism. This application featuring a very complete tourist profile; however, that profile is not used for any kind of serious reasoning, such as a RS.

Since the data available is not very extent, it was not chosen to perform a formal comparison between the different systems. Still, informal considerations can be extrapolated. In a very broad statement, the current main flaw regarding tourism RSs is the *poor UM* backing them up. Most systems rest their efficiency on a single UM technique; even if such modeling is not incorrect, it is certainly not enough, considering the complexity of the human being in various aspects. Another approach surfacing in the latest years is the overrated preference for social filtering methods [19]. While the use of this kind of recommendations is not uninteresting (we also embrace them in this very work), it is clear that the user itself is still the most important and primary source of recommending material, one whose deep analysis has not yet been performed in such a complete manner as will be presented. The growth of social networks and the current Web 2.0 is making RS *too social-targeted*, providing the UM architecture with a void that can still be exploited. To finish, and in such a domain where both personal and social interests are at stake, it is with surprise that systems do not rely on information related to the *psychological / behavioral* side of users, a key component of decision-making processes.

Given this action space for improvement, the proposed work tries to evolve the filtering of RS and the way systems can benefit from different kinds of user information, by using a complex UM framework. In the next pages, a RS architecture will be presented.

### 3 Development

#### 3.1 Recommender System

Although tourism systems are more than a mere RS, that is the most important step regarding the user point of view, since the underlying mechanisms are invisible to them. Contrary with what the state of the art may make us think, adapting the system to accommodate user's preferences and needs can be accomplished using a variety of techniques. Figure 1 presents all techniques contemplated in the devised RS, as well as their importance.

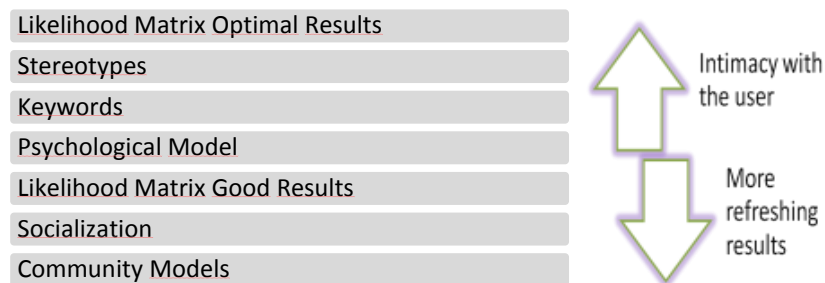


Fig. 1. Recommender System's Techniques

Given the internal nature of techniques (Fig 1) it is clear that, for example, socialization techniques are less important than stereotypes, since the assumptions taken for granted socially might be completely wrong, a situation with much less probability in the other technique. The criteria used for such an order was the absolute relation between its results their intimacy with the user, also according with some tests executed. Another interesting aspect that can be experienced is that, in opposition with this intimacy order, lower techniques offer a much higher probability of giving refreshing results. Although out of the scope results are considered by many authors as a bad subject, promoting over-specialization, in this work we do not abide by those opinions. In fact, we express that such a point of view is much necessary in order to evolve user tastes, showing a sign of intelligence that users can interpret. Table 1 relates the presented techniques with the classical literature approaches. The proposed system makes a commitment into balancing (1) the acceptance of traditional techniques and (2) introducing innovations into each one of them and also proposing a new one [9][10][15][16][17].

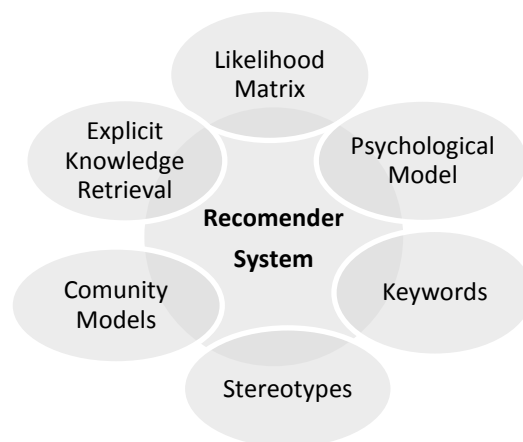
**Table 1** - Comparison Between Literature Techniques and Proposed System

<b>System Technique</b>	<b>Literature Technique</b>
Likelihood Matrix	Knowledge-based Filtering
Keywords	Content-based Filtering
Socialization	Collaborative Filtering
Stereotypes and Psychological Model	Behavioral / Psychological Filtering

All of the above techniques will be explained regarding their internal mode of operation and evolution. In fact, those mechanisms are independent of the RS, being able to function without this component on top of them. Following is an explanation of how these play their role, in fact, in order to help the RS.

### 3.2 User Modeling Techniques

As it was earlier implied, a powerful RS is endorsed by several techniques that maximize the accuracy of the necessary user information in a variety of ways. Our UM methodology (which backs up the RS) must therefore be a collaborative effort of several sub-systems, each of them responsible for the retrieval of part of user data. Furthermore, it is also believed that knowing a certain user information space by using more than one method simultaneously successfully increases confidence in existent assumptions and divides responsibility amongst various techniques, which ultimately results in a system with more solutions and contingency plans [2][3][4]. This knowledge retrieval spirit represents the true evolution of our system against the considered state of the art. The system's knowledge discovery components can be observed in Fig. 2.



**Fig. 2.** Knowledge Discovery Mechanisms' Architecture

It can easily be seen that tourism applications' upper-level core functions (namely the RS) gather information throughout all sub-systems and merge that data into a coherent user profile, in order to generate new information. These representation formalisms represent an advanced view of the user profile and allow for the generation of value added (possibly new) knowledge.

### 3.2.1 Likelihood Matrix

The likelihood matrix is responsible for linking the user with each one of the categories created in the POI taxonomy, being classified as a linear model in what literature techniques are concerned. It ranges from -1 to 1, where -1 means total unlikelihood and 1 represents complete interest. The use of likelihoods between types of POI and the user is not new, as it has already been used in other systems and using different formats [2]. However, the techniques employed here allow rarer analysis such as positive / negative likelihoods, thus the choice for the -1 to 1 floating number. Plus, the underlying taxonomy is much richer and structured than the majority of other tourism systems [19]. This mechanism is the basis of the stereotype module and thus both components work together in order to provide an over-confident representation of user interests. Although one technique is based on the other, their abstraction level is different, triggering different results by both components. The proposed approach coherently represents both user likes and dislikes, by maintaining a negative and positive action space in which assumptions can diverge within. Moreover, by defining optimal thresholds, it is possible to identify the most important POI categories within users and trigger adequate response. The likelihood matrix is fed by Application Interaction Triggers (AITs), which monitor important and relevant actions executed by users within activity sessions..

### 3.2.2 Stereotypes

Stereotypes within the tourism domain are not inexistent, but they do not provide evolutionary capabilities as those that were developed and employed here, although they are described in other interesting manners, such as demographic attributes [7]. Stereotypes represent a widely used information abstraction mechanism used to group users into categories. The work made in [7] was fundamental in our stereotype component definition, as several ideas were mirrored in our system. Our stereotype system can be explained through a set of development guidelines which originated it. First of all, the POI taxonomy was *re-conceptualized* into hierarchical terms which would better serve as the basis for the stereotype construction; then, an initial set of *stereotypes* was created, each of them being fully described using the previous terms which will form the comparison basis for that stereotype to be linked to a user. Finally, *mechanisms* were created to compensate for an eventual insufficiency that might describe the initial set of stereotypes, as well as the suitability of their terms.

The concept hierarchy's creation process resulted in adequate terms to be used in the stereotype creation. An example of a mapping between the original taxonomy and these concepts is the following: Pubs + Nightlife + Festivities = Partying. Six initial stereotypes were created, along with the activation conditions to trigger the link

between a user and a particular stereotype. Upon new user arrival, the registration form will be responsible for matching him / her to one of the existing stereotypes within the system. From that point on, different user actions will put him into different stereotypes, which will trigger different approaches in several system components, such as the RS [7]. The activated stereotypes for a given user in a certain point in time will always be those whose value surpasses the Activation Threshold, which will determine the nature of recommendations given to that user. A very important feature of stereotypes is, as said before, the group of mechanisms for coping with stereotype decay that were defined. Those are: (1) propose underused stereotypes for removal, (2) propose underused stereotype conditions for removal, (3) propose overused conditions not included in stereotypes and finally (4) propose new stereotypes based on user profiles (eliminates the grey sheep individual issue).

It's in these four processes that stereotypes can obviously be inserted into the category of clustering and classification techniques [20].

### 3.2.3 Keywords

The concept behind keywords is socially very powerful, addressing knowledge about items in a fashion more intimate to the user. Plus, by employing a knowledge representation with significant value to people, it also embraces the social power of the system. Keywords require significant user participation in order to be fully profitable; however, in recent years, due to the social web appearance, the idea got more doable [8]. The advantages of keywords are numerous:

- The power of keywords surpasses any other kind of controlled information representation, like the POIs taxonomy, granting value-added knowledge;
- It means the featuring of items in the user point-of-view, which enhances the application's real world / social sense and brings it closer to the user;
- Slowly and partially remove the importance of system-defined approaches (more strict and controlled), putting the power of the application in the users hands;
- By storing and relating keywords with users, tags can propose themselves as yet another means of filtering items to the user, along with all other methods.

Several approaches for initially inserted tags within the items were devised, without having to take an intensive cognitive process of cataloging them. By automatically gifting items with tags, one of the few downsides of this kind of social media, the cold-start problem, can be diminished. The following types of tags are initially and automatically set for POI:

1. Keywords that relate to the POIs class which classifies the POI. For instance, a zoo would have the tags *cultural*, *natural\_parks* and *animal\_preserves* automatically associated upon creation;
2. Keywords that pertain to all features related with that POI, which were previously added to the respective POIs class. For example, a certain restaurant might have the keywords *spicy* and *Mexican*;
3. Keywords that pertain to special words found within the name and description of the item, by using text mining techniques. For example, the bridge "D. Luis I" might have the keywords *bridge*, *Gustave\_Eiffel* and *iron*.

### 3.2.4 Psychological Model

The user psychological model is in constant evolution, as the user interacts with the system and gives it traces of his personality evolution. This information might also be initially given by the user, to propel the system with a more coherent start. The features selected to be part of this module were based on psychological models devised by authors along the years [5][12][13]. Four attributes were elected: liveliness, perfectionism, outdooriness and creativity. They range from 0 to 1, representing the two extremes of that feature. The user psychological evolution proceeds as follows: each POI category is labeled by one or more of the four psychological attributes; then, interactions with POIs or POIs categories will trigger analysis between psychological models of both the user and the corresponding POIs classes. The analysis results will feed and evolve the user behavioral model towards the comparison o object (the POIs class), therefore adjusting it and changing the input of all components depending on it, such as the RS.

The effective and useful use of psychological data is very rare in the current computational scene in general, let alone in the tourism or RS [10]. The user psychological model, along with the use of stereotypes, represents a new approach - *behavioral-based* - in modeling and recommending items to users and contributes to the innovative nature of this work. This technique is also an example of a predictive statistical linear model, since recommendations are based on top of past experiences.

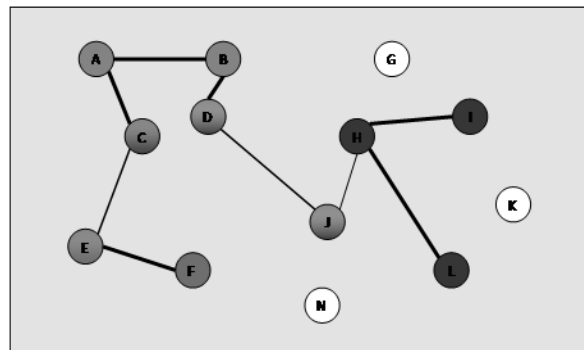
### 3.2.5 Socialization

The recommendation of items based on user's social data has been greatly used in the past years. However, as was already explained, it occupies the sixth position within the order of the filtering techniques, due to the criteria chosen for that order. In fact, one of the main objectives of the devised project is the exploitation of intra-user filtering techniques, given the space for improvement. Even so, this technique presents innovations of its own, profiting from the complex UM architecture that the system uses. Indeed, the manner by which users are matched similar against the main persona is made much more coherent, by using the last four explained techniques. This way, despite the technique being given less importance, it presents, in fact, a much more sustained theory for matching users, in respect with the current state of the art. Actually, the same comparison method is also used in a friend RS existent within the yet superficial social network which gifts the system.

### 3.2.6 Community Models

Our UM architecture makes use of a group of two Community Models (CMs). CMs are an adaptation of Neural Networks (NN), specifically a rather simple form of those. CMs generate two-dimensional representations of the data previously fed into its mechanisms [6], evolving and activating its nodes and links. CM's output is mirrored by other techniques, such as dynamic neural networks and Bayesian networks. However, the amount of increased work necessary to deploy those kinds of techniques compared to the value-added knowledge they provide, as opposed to these CMs, was not profitable [20]. CMs, when properly fed, can give us interesting patterns and

associations between items that were previously invisible. Both CMs represent POIs in their nodes, while the relations between nodes represent POIs co-occurrence. The first CM is about system navigation (user sessions, *clickstream* analysis, etc.), while the second is concerned about effective POI commitment, such as the generation of a visiting route. Fig. 3 displays a fictional portion of one of the CMs. The color of each node is a reference to geographical position, meaning that CMs might act as a clustering technique, naturally grouping POIs, physically related or not.



**Fig. 3.** Community Model's Example

Although much of the presented CM's usefulness is already attended in current systems, some of them are unprecedented, given the current state of the art:

1. Discover POI associations, by analyzing patterns contained within selected groups of items. Found associations will, apart from triggering important strategic decisions by related entities, become an effective help in understanding user personalities and also improving the RS's efficiency. This kind of operation is a very simple but effective kind of Rule Association (RA) analysis;
2. If, ideally, the most used POI would be the same that were previously viewed, that might not always be the case. By searching for abnormalities between viewed versus used items, mechanisms can be triggered, which may include an increase in system visibility of some least selected items, for instance.

## 4 Evaluation

In this section, some evaluation guidelines will be presented in order to demonstrate the capabilities of the devised system, as well as a survey that targeted the so-far users of the application.

*Startup quality:* the devised prototype delivers an increased startup quality of response concerning filtering features and personalization mechanisms. By making use of a clever and abstracting model for initially asked information, the application can instantly achieve a coherent user profile and subsequent sustained suggestions. It can, technically, fill half of the user model information within the initial form. This startup quality is not performed by other systems [18].



*Transparency:* besides profiting from an automatic UM platform which does everything in an automatic fashion, the user can also be invited into viewing, in a transparent manner, everything that the system believes about him. By making use of a friendly interface, users can increase confidence of such critical information and enhance RS's results immediately.

*Recommender System:* the RS state of the art (see [9][10][15][16]) has reached a very critical state regarding innovation. The UM platform here presented forms a very diverse basis for RS computation and introduces a new way of filtering complex-domain items: behavior-based. This technique merging causes the RS to output items with diverse sources and assumptions, increasing likelihood for item success. Plus, by making use of several techniques, the RS has the ability to overcome problems that certain techniques might have, such as under-confidence or the cold start problem.

*On-the-fly profile evolution:* almost all of the UM building blocks are propagated immediately, causing the consequent RS to always output fresh results. Thus, without damaging any interaction flow during user sessions, the user is immediately gifted with up-to-date responses from any application-level process. Such instant adequacy of results is not performed, for example, by the TripAdvisor application.

*Diverse knowledge:* the UM components that comprise the system use knowledge representation formalisms of diverse sources. One of the most important ways of analyzing these different sources is through their degree of control. While controlled knowledge (like the likelihood matrix) allows for sustained and guaranteed outputs, uncontrolled knowledge, existent in keywords, for example, grants users freedom and control in evolving the way the system works.

*Survey:* regarding the user survey that was made, the following conclusions were taken: (1) users were pleased by the short initial form; (2) users enjoyed the transparent spirit of the system and show interest in using the User Area; (3) users have found the RS's results satisfactory, lacking only a higher degree of speed, due to the complex reasoning tasks that are put to work when the RS fires off.

## 5 Conclusions

Machine learning reasoning hasn't been fully explored in the tourism domain, particularly in what the tourist model is concerned [20]. Tourism applications are indeed in a widespread state, but they still haven't taken the evolution step of intelligence, as well as usability. Users still have to perform several actions to achieve what they really want, and a substantial amount of smart reasoning from the system doesn't seem still present [18][19]. Moreover, user information has been misused and falsely utilized. On one hand, current systems ask more information than they really use, which is a downside both usability-wise and resource-wise [19]. On the other hand, UM and user adaptive-related mechanisms are used in a very poor manner, based on very weak assumptions. In the proposed system, a modeling platform that requires only a minimal amount of user effort and still manages to create a very capable user image is deployed. Next is a list of advantages that, when compared with current projects (see section 2), dictate the importance and innovative nature of this project: (1) advanced and innovative UM, (2) on-the-fly user profile update, (3) hot-start results quality, (4) behavioral-filtering introduction and (5) multi-technique and

heterogeneous RS. From our point of view, the user model present in this paper (stage one), along with the more complex knowledge inference mechanisms that constitute the devised system (stage two) are two excellent basis for any tourism-focused system. The architecture presented can actually be applied to completely different scenarios and domains, given some content minor content adjustments.

## References

1. Gay G. & Hembrooke, H.: Activity-Centered Design: an Ecological Approach to Designing Smart Tools and Usable Systems. MIT Press (2004)
2. Fink J. & Kobsa A.: User Modeling for Personalized City Tours. In: Artificial Intelligence Review. Volume 18, Number 1. Kluwer Academic Publishers (2002)
3. Kobsa A. Generic User Modeling Systems. In: User Modeling and User-Adapted Interaction. Volume 11, Issue 1-2. Kluwer Academic Publishers (2001)
4. Kobsa A.: User Modeling: Recent Work, Prospects and Hazards. In: Adaptive User Interfaces: Principles and Practice. North Holland Elsevier (1993)
5. Jung, C. G.: Psychological Types. Princeton University Press (1971)
6. Jennings A. & Higuchi H.: A Personal News Service Based on a User Model Neural Network. In: User Modeling and User-Adapted Interaction. Volume 3, Issue 1. Springer Netherlands (1992)
7. Rich E.: User Modeling Via Stereotypes. In: Cognitive Science: A Multidisciplinary Journal. Volume 3. Morgan Kaufmann Publishers Inc. (1979)
8. Mathes A.: Folksonomies - Cooperative Classification and Communication through Shared Metadata. In: Computer Mediated Communication - LIS590CMC (2004)
9. Berka T. & Plößner M.: Designing Recommender Systems for Tourism. In: ENTER 2004 (2003)
10. Felfernig A., Gordea S., Jannach D., Teppan E. & Zanker M.: A Short Survey of Recommendation Technologies in Travel and Tourism. In: ÖGAI Journal. Volume 25, Issue 2 (2006)
11. Coelho B., Martins C. & Almeida A.: ATOMIC - Arquitetura para Turistas com vista à Organização, Modelação e Inferência de CONhecimento - Representação do Utilizador. In: IADIS Ibero-Americana - WWW / Internet 2008 (2008)
12. Oliver P. J. & Srivastava S.: The Big-Five Trait Taxonomy: History, Measurement and Theoretical. In: Handbook of Personality: Theory and Research. Guilford Press (1999)
13. Cattell, R. B.: Handbook for the Sixteen Personality Factor Questionnaire. The 16 P.F. Test - Forms A, B, and C. Institute for Personality and Ability Testing (1962)
14. FilmTrust, <http://trust.mindswap.org/FilmTrust/> (2010)
15. Burke R.: Knowledge-Based Recommender Systems (1999)
16. Pazzani M. J. & Billsus D.: Content-Based Recommendation Systems. In: Lecture Notes in Computer Science. Volume 4321. Springer Berlin / Heidelberg (2007)
17. Porter J.: Watch and Learn: How Recommendation Systems are Redefining the Web (2006)
18. TripAdvisor, <http://www.tripadvisor.com/> (2010)
19. WAYN, <http://www.wayn.com> (2010)
20. Zukerman I. & Albrecht D. W.: Predictive Statistical Models for User Modeling. In: User Modeling and User-Adapted Interaction. Volume 11, Numbers 1-2, Pages 5-18 (2000)