Resource Recommendation for Intelligent Environments Based on a Multi-Aspect Metric

Aitor Almeida[1] , Eduardo Castillejo[1], Diego López-de-Ipiña[1], Marcos Sacristán[2], Javier Diego[3]

[1]DeustoTech –Deusto Institute of Technology, Universidad de Deusto Avenida de las Universidades 24, 48007, Bilbao (Spain)
{aitor.almeida, eduardo.castillejo, dipina}@deusto.es

[2]Treelogic Parque Tecnológico de Asturias, 30, E33428, Llanera (Spain)
marcos.sacristan@treelogic.com

[3]Logica Avda. Manoteras, 32, Madrid 28050 (Spain)
javier.diego@logica.com

Abstract. Intelligent environments offer information filled spaces. When trying to navigate among all the offered resources users can be overwhelmed. This problem is increased by the heterogeneous nature of resources in smart environments. Users must choose between a plethora of services, multimedia information, interaction modalities and devices. But at the same time the unique characteristics of smart spaces offers us more opportunities to filter these resources. To help users find the resource that they want and need we have designed a multi-aspect recommendation system that takes into account not only the features of the resource and the user, but also context data like the location and current activity. The developed system is flexible enough to be applied to different resource types and scenarios. In this paper we will describe the identified aspects and how they are merged into a single metric.

Keywords. Intelligent Environments, resource recommendation, context aware, accessibility

1 Introduction

Intelligent Environments host a diverse ecosystem of devices, services and multimedia content. Users interact with these resources, either by using them directly or consuming them via a plethora of mobile devices. As the environments become more sophisticated even more of these resources will be made available for the user. All these resources can be overwhelming, making it difficult to find those more suitable for the current situation. In order to tackle this problem Intelligent Environments must be able to react to user needs in order to fulfill the request and desires of the users. To do this the system must know the user preferences, tastes and limitations. It must be capable of analyzing the different aspects that define a resource to offer the most suitable one.

To do this the recommendation system must be able to process the heterogeneity of the analyzed resources. To be able to do this recommendation we have identified the aspects of a resource that can be used to describe it in a Intelligent Environment. These aspects take into account both the resource and user features and the current context. Our recommendation system approach has several advantages: 1) It is applicable to any type of resource; 2) We evaluate different aspects of the resource taking into account the characteristics of the content, the needs and capabilities of the user and data from the current context. This allows us to create a comprehensive picture of the current situation to recommend the most suitable resource; 3) The process can be configured by modifying the weight of each individual aspect in the final metric. This allows us to adapt the recommendation system to specific domains and 4) Our system not only analyzes the current situation of the user, it also takes into account what his next actions can be to anticipate future needs.

2 Related Work

Since the mid-1990s recommender systems have become an important research area attracting the attention of e-commerce companies. Amazon [1], Netflix and Yahoo! Music [2] are widespread examples on making recommendations to its users based on their tastes and previous purchases. Although these systems have evolved becoming more accurate, the main problem is still out there: to estimate the rating of an item which has not been seen by users. This estimation is usually based on the rest of items rated by the current user or on the ratings given by others where the rating pattern is similar to the user's one. Therefore, the problem consists on extrapolating somehow the utility function (which measures the usefulness of an item to a user) to the whole rating space. This utility function is represented by all the ratings made by the user. This way, recommendation engines have to be able to predict or estimate the ratings of the not yet rated items for users. Content-based systems recommend items which are similar to those that a user rated positively in the past [3]. Shardanand et al [4] state some of the problems of this approach, as the vagueness in the description of an item, which clearly affects the whole system. Items need to have enough descriptive features to enable the recommendation engine to recommend them accurately. The problem is that different items with the same features can be indistinguishable to the system. Collaborative filtering techniques deal with the concept of similarity between users. The utility of an item is predicted by those items which have been rated by similar users. Sarwar et al [5] defend this approach by defining collaborative filtering as the most successful recommendation technique to date. In [4] a personalized music recommendation system is presented, namely Ringo, which is a social information filtering system which purpose is to advise users about music albums they might be interested in. By building a profile for each user based on their ratings, it identifies similar users so that it can predict if a not yet rated artist/album may be to user's liking. LikeMinds [6] defines a closeness function based on the ratings for similar items from different users to estimate the rating of these items for a specific user. It considers a user which has not already rated the item and a so-called mentor who did it. Introducing two new concepts (horting and predictability) horting is a graph-based technique in which users are represented as nodes and the edges between them indicate their similarity (predictability) [7]. The idea is similar to nearest neighbor, but it differs from it as it explores transitive relationships between users who have rated the item in question and those who have not. In order to reduce the limitations of previously reviewed methods, hybrid approaches combine both of them [8]. Others have introduced new concepts to this area, such as semantics and context [9].

However, one of the most important improvements in the recommendation systems field is the definition of measures (or aspects) to describe the utility and relevance of the items. Aspects play an important role in data mining, regardless of the kind of patterns being mined [10]. Users' ratings are a good way to trace the interestingness and the relevance of items. Despite of the ratings, there are many measures which allow us to go into these items taking into account the use of them (their consumption) by the users. In other words, we look into the behavior of users for measuring their interestingness for these "items" (for now on we will refer items as resources). From our point of view a resource could be a product, an application or any kind of service (e.g., multimedia, news and weather or connectivity infrastructure services). We have studied several measures from the literature to evaluate those which best fit in our recommender system, such as minimality [11,12], reliability [13], novelty [14], horting, predictability and closeness [10], and utility [5].

3 Resource Evaluation Metric

To be able to evaluate the suitability of the resources for a given user we have identified a series of aspects that define any given resource. These aspects must be generic enough to be able to use them to describe any type of resource (services, content and so on) and expressive enough to capture the different facets of the resources. In the current implementation (see Figure 1) we have considered four of them, but we discuss the other ones in the future work section. The four aspects that we currently take into account are the following: Predictability, Accessibility, Relevancy and Offensiveness. Each one of those aspects is used in the calculation of the suitability value (see Formula 1). The weight of each aspect on the final value can be modified to better adapt the recommendation system to the specific domain of each smart environment (e.g. to the business plan of a hotel, to prioritize those aspects demanded by the clients). The suitability value is always personalized to a specific user and can change over the time along the preferences of the user.

$$M_{tot} = \sum \omega_i f_i \quad (1)$$

Where M_{tot} is the value of the suitability of each resource, ω_i is the weight for an aspect and f_i is the value of the aspect of a resource. The values of the aspects are normalized.

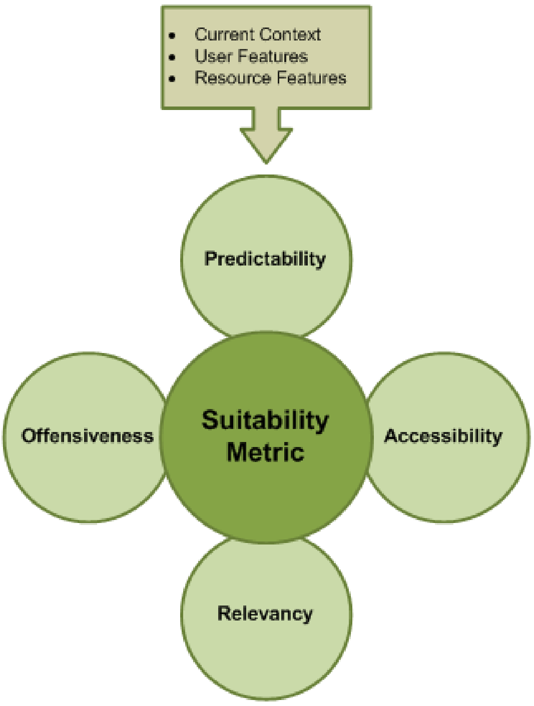

Fig. 1. The multi-aspect metric

3.1 Predictability

The first aspect we evaluate is the predictability. This aspect reflects how likely a resource is to be used based on the resources consumed previously. This likeliness is expressed as a probability value between 0 and 1. We use Markov Chains to create the model of the user's resource usage. This model allows us to ascertain patterns in the user behavior. E.g. When one user stays on the hotel his morning routine consists in using the "Press Digest" to recover the headlines of the day, the "Room Service" to order breakfast and the "Transport Service" to call a taxi. With the generated model we will able to predict that after using the "Room Service" the most probable service to be consumed is the "Transport Service". To build the transition matrix for the Markov Chains we use the previous history of the user's resource consumption as the training set. This transition matrix can be retrained with the new data recovered from the user with each visit to the hotel, adapting itself to the changes in the user preferences. As we discuss in the future work section one of the main problems with using Markov Chains is that we only take into account the last consumed resource to predict the next one due to the Markov Property.

3.2 Accesibility

One of the most important aspects is the accessibility features of the resource. Users of intelligent environments possess a wide variety of abilities (sensorial, cognitive and so on) that must be taken into account to assess the suitability of the resources. Whatever the resource is, users must be able to consume it. We have used the user abilities taxonomy proposed in [15]. We have restricted the user abilities to three groups: 1) *Sensorial abilities*: Those abilities related to the user input; 2) *Communicational* abilities: Those abilities related to the user output and 3) *Physical*: Those abilities related with the capability of the user to move his extremities .

Each resource has two types of abilities associated, the required and recommended user abilities. If the user does not have one of the required abilities the value of the aspect is automatically set to 0. This is done to reflect the fact that the user can not consume the resource, thus being completely useless for that user. If the user does not have a recommended ability the accessibility value receives a penalization (see Formula 2).

$$A_{acc} = 1 - \omega|Rec_{not}| \tag{2}$$

Where A_{acc} is the accessibility value for the resource, ω is the penalization weight and $|Rec_{not}|$ is the number of recommended abilities not met by the user.

3.3 Relevancy

This aspect measures the importance of a given resource to the user's current context. For example, a user jogging may be interested in the location of parks and running routes but a user having breakfast in the hotel may be interested instead in the public transports available in the city. One of the main problems we encountered evaluating this aspect was the selection of the context variables. The selected variables must be significant enough to be applicable to any type of resource in any given domain. We have identified three context variables that meet these requisites: 1) *User location.* In the tourism domain we have considered the following locations: client's room, hotel's lobby, hotel's restaurant, hotel's swimming pool, hotel's gymnasium and outside the hotel; 2) *Time of the day*. We have divided the day in twelve periods of two hours and 3) *Current activity.* In the tourism domain we have identified seven activities: sleeping, morning routine, having breakfast, exercising, working, shopping and visiting tourist attractions.

The context information is provided by other modules of the THOFU project that are out of the scope of this paper. Using the usage data recollected from the users we have trained a soft classifier that, given those three context variables, calculates the relevancy of a resource. For the classifier we have used a nearest neighbor search. To implement this classifier we have used the libraries included in the Weka framework. We have used LinearNNSearch as the nearest neighbor search algorithm, with the Euclidean distance as the distance function.

3.4 Offensiveness

This aspect measures the suitability of a resource based on a rating system. We use the age categories (3, 7, 12, 16 and 18) and the content descriptions (violence, bad language, fear, sex, drugs, gambling, discrimination and online) developed for the PEGI (Pan European Game Information) rating system. To evaluate it we use a similar system that the one used in Section 3.1 to calculate the accessibility, but taking the age categories as required constraints and the content descriptions as the recommended ones.

4 Use Case

To better illustrate how the developed system works we will explain how the system works taking two different users as examples. The first user is a 27 year old male with a hearing impairment. The second one is a 6 year old child. The users have five resources available to them in this example: The wake up service (R1), the room service (R2), the press digest (R3), the multimedia system (R4) and the transport service (R5). For this example the weights for the metric calculation are:

- *predictability* and *relevancy* have a weight of 1
- *accessibility* and *offensiveness* have a weigh of 0.5

We assume that both users are in their rooms and that the wake up service has just been activated by an alarm.. The wake up service and multimedia system both have hearing requirements, but offer alternative means to use them. The first user has not stated any content restriction. The results are shown in Table I.

Table 1. Results for the first user

	Predictability	Accessibility	Offensiveness	Relevancy
R1	0.10	0.9	1	0.8
R2	0.60	1	1	0.7
R3	0.30	1	1	0.4
R4	0	0.9	1	0.2
R5	0	1	1	0.3

The second user has not any disability, so every resource attains the maximum score in accessibility. The press digest has a minimum age category of 7 and it receives a score of 0 in offensiveness. The results are shown in Table II.

Table 2. Results for the second user

	Predictability	Accessibility	Offensiveness	Relevancy
R1	0.45	1	1	0.2
R2	0.05	1	1	0.1
R3	0	1	0	0.1
R4	0.50	1	1	0.9

R5	0	1	1	0

Using the Formula 1 the recommended resource for the first user will be the room service (R2) in this scenario.

$$M_{tot} = 1 \times 0.6 + 0.5 \times 1 + 0.5 \times 1 + 1 \times 0.7 \quad (3)$$

In the case of the second user the selected resource will be the multimedia system (R4).

$$M_{tot} = 1 \times 0.5 + 0.5 \times 1 + 0.5 \times 1 + 1 \times 0.9 \quad (4)$$

5 Conclusion and future work

The number of available resources in smart environments can be overwhelming. User can consume a large number of services and multimedia content. In order to tackle this problem we have described a resource recommendation system based on a multi-aspect metric. This recommendation system is specially tailored for intelligent environments, taking into account the user context. Our approach has several advantages: 1) Our recommendation mechanism can be applied to any resource (services, multimedia data, interaction mechanisms…) in an smart environments. This is due to its multi-aspect nature and its configurability; 2) We provide a holistic approach to the recommendation problem, taking into account multiple variables to achieve the best possible recommendation; 3) Finally the metric can be specifically adapted for different domains tailoring the weights of each aspect. This allows us to create specific solutions for each problem.

One of the problems identified in this approach is the use of Markov Chains to evaluate the predictability aspect. With the use of Markov Chains we only evaluate the current event and not the previous events that preceded it. In order to tackle this problem we plan to explore the use of time series to improve the forecasting algorithm.

We are also analyzing a more extensive set of aspects that will give us a better picture of the evaluated resources. We are currently studying the inclusion of the following aspects:

- *Timeliness*: evaluates how up to date is the information of a resource.
- *Satisfaction*: measures the opinion of the users about a resource.
- *Attention*: The average number of interactions per time unit with a consumed resource.
- *Closeness*: Evaluates what resources are consumed by similar users.

Adding these new aspects we aim to create more significant resource recommendations that meet better the user's needs. Finally we would like to include in the context data information about the vagueness and uncertainty of the model. This will allow us to model the context more realistically and will improve the overall preciseness of the system.

Acknowledgment.
This work has been supported by project grant CEN-20101019 (THOFU), funded by the Spanish Centro para el Desarrollo Tecnológico Industrial (CDTI) and supported by the Spanish Ministerio de Ciencia e Innovación.

References

1. G. Linden, B. Smith, and J. York, "Amazon. com recommendations: Item-to-item collaborative filtering," Internet Computing, IEEE, vol. 7, no. 1, pp. 76–80, 2003.
2. P. L. Chen et al., "A Linear Ensemble of Individual and Blended Models for Music Rating Prediction."
3. D. M. Pennock, E. Horvitz, S. Lawrence, and C. L. Giles, "Collaborative filtering by personality diagnosis: A hybrid memory-and model-based approach," in Proceedings of the 16th conference on uncertainty in artificial intelligence, 2000, pp. 473–480.
4. U. Shardanand and P. Maes, "Social Information Filtering: Algorithms for Automating 'Word of Mouth'," 1995.
5. B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, "Item-based collaborative filtering recommendation algorithms," in Proceedings of the 10th international conference on World Wide Web, 2001, pp. 285–295.
6. D. Greening, "Building consumer trust with accurate product recommendations," LikeMinds White Paper LMWSWP-210-6966, 1997.
7. C. C. Aggarwal, J. L. Wolf, K. L. Wu, and P. S. Yu, "Horting hatches an egg: A new graph-theoretic approach to collaborative filtering," in Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, 1999, pp. 201–212.
8. M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin, "Combining content-based and collaborative filters in an online newspaper," in Proceedings of ACM SIGIR Workshop on Recommender Systems, 1999, p. 60.
9. S. Kim and J. Kwon, "Effective context-aware recommendation on the semantic web," International Journal of Computer Science and Network Security, vol. 7, no. 8, pp. 154–159, 2007.
10. L. Geng and H. J. Hamilton, "Interestingness measures for data mining: A survey," ACM Computing Surveys (CSUR), vol. 38, no. 3, p. 9, 2006.
11. B. Padmanabhan and A. Tuzhilin, "Small is beautiful: discovering the minimal set of unexpected patterns," in Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, 2000, pp. 54–63.
12. Y. Bastide, N. Pasquier, R. Taouil, G. Stumme, and L. Lakhal, "Mining minimal non-redundant association rules using frequent closed itemsets," Computational Logic—CL 2000, pp. 972–986, 2000.
13. P. N. Tan, V. Kumar, and J. Srivastava, "Selecting the right interestingness measure for association patterns," in Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, 2002, pp. 32–41.
14. S. Sahar, "Interestingness via what is not interesting," in Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, 1999, pp. 332–336.
15. Aitor Almeida, Pablo Orduña, Eduardo Castillejo, Diego López-de-Ipiña, Marcos Sacristán. Imhotep: an approach to user and device conscious mobile applications. Personal and Ubiquitous Computing. Volume 15 Issue 4, pp 419–429 January 2011.