A Novel Strategy for Recommending Multimedia Objects and its Application in the Cultural Heritage Domain

Massimiliano Albanese, George Mason University, USA

Antonio d'Acierno, ISA, National Research Council, Italy

Vincenzo Moscato, University of Naples, Italy

Fabio Persia, University of Naples, Italy

Antonio Picariello, University of Naples, Italy

ABSTRACT

One of the most important challenges in the information access field, especially for multimedia repositories, is information overload. To cope with this problem, in this paper, the authors present a strategy for a recommender system that computes customized recommendations for users' accessing multimedia collections, using semantic contents and low-level features of multimedia objects, past behaviour of individual users, and social behaviour of the users' community as a whole. The authors implement their strategy in a recommender prototype for browsing image digital libraries in the Cultural Heritage domain. They then investigate the effectiveness of the proposed approach, based on the users' satisfaction. The preliminary experimental results show that the approach is promising and encourages further research in this direction.

Keywords: *Information Overload, Multimedia Browsing, Multimedia Information Retrieval, Multimedia Ranking, Recommender Systems*

1. INTRODUCTION

Multimedia data allow fast and effective communication and sharing of information about people lives, their behaviors, work, interests, but they also are the digital testimony of facts, objects, and locations. The widespread availability of cheap media technologies (e.g., digital and video cameras, MP3 players, and smart phones) dramatically increased the availability of multimedia data. Images and videos are used, by media companies as well as the public at large, to record daily events, to report local, national, and international news, to enrich and emphasize web content, as well

DOI: 10.4018/jmdem.2011100101

as to promote cultural heritage. Furthermore, through digitization of all types of data and records, multimedia data plays an increasingly critical role in government administration, from security and justice to the health system.

As a result, huge data collections, in the form of digital video and image libraries, digital documents, news archives, shopping catalogs, virtual museums, and so on, are widely available, determining the well-known problem of *information overload*. From such a massive amount of data, it is very difficult for a common user to obtain her/his preferred ones, and despite the great amount of research work done in the last decade, retrieving and suggesting information of interest from very large repositories still remains an open issue, especially in the case of multimedia collections.

To cope with this problem, a number of algorithms and tools – generally referred to as *Recommender Systems* – are being proposed to facilitate browsing of large data repositories, and thus to realize the transition from the "era of search" to the "era of discovery". Recommender systems help people in retrieving information that match their preferences by recommending products or services from large number of candidates and support people in making decision in various contexts: what items to buy (Zhang & Wang, 2005), which movie to watch (Qin et al., 2010) or even who they can invite to their social networks (Kazienko & Musial, 2006).

They are especially useful in the environments with a vast amount of information where it is difficult to express the semantics of a query since they allow an automatic selection of a small subset of items that appears to fit to the user needs (Adomavicius & Tuzhilin, 2005). The main problem for multimedia is that the *semantic gap* between users and contents is sometimes so large that very little previous work succeeds in building an effective multimedia recommender system.

In such a context, the main goal of this work is to present a novel approach to recommendation for multimedia objects, based on an "importance ranking" algorithm that strongly resembles the well known *PageRank* ranking strategy (Albanese et al., 2011). We propose a method that computes customized recommendations by originally combining intrinsic features of multimedia objects (low level and semantic similarities), past behavior of individual users and overall behavior of the entire community of users. Eventually, we have implemented the proposed strategy in a software prototype for browsing digital libraries related to a famous on-line collection of paintings and measured its effectiveness with respect to a user-centric evaluation.

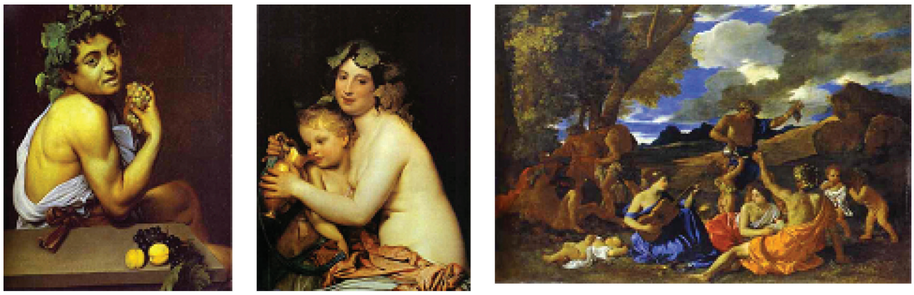
The paper is organized as follows. Section 2 presents some motivating examples that justify the utility our work in different contexts related to Cultural Heritage domain. Section 3 discusses the state of the art of recommender systems, including those applied in the multimedia realm. Section 4 shortly describes the proposed strategy for recommending multimedia objects. Section 4 illustrates the system architecture and provides some implementation details. Section 6 reports preliminary experimental results; finally, Section 7 gives some concluding remarks and discusses future work.

2. MOTIVATING EXAMPLES

In this section we report two typical scenarios in the Cultural Heritage domain that aim at showing how a recommendation system could desirably work during both a virtual – but also a real – visit of an art gallery.

First, let us consider an on-line art museum offering web-based access to a multimedia collection of digital reproductions of paintings. For instance, let us consider users visiting such a virtual museum and suppose that they request, at the beginning of their tour, some paintings depicting the "Bacchus" subject. While observing such paintings, they are attracted, for example by Caravaggio's painting entitled "Self-Portrait as Sick Bacchus" (Figure 1a). It would be helpful if the system could learn the preferences of the users, based on these first interactions and predict their future needs by suggesting other paintings representing the

Figure 1. Paintings depicting "Bacchus"



a b c

same or related subjects, depicted by the same or other related authors or items that have been requested by users with similar preferences, thus avoiding the use of a classical keyword-based search engine.

As an example, a user who is currently watching the Carvaggio's painting in Figure 1a might be recommended to see the Feodor Bruni's painting entitled "Bacchante Giving Wine to Cupid" (Figure 1b), that is quite similar to the current picture in terms of color, shape and texture, and "Andrians or The Great Bacchanal with Woman Playing a Lute" by Nicolas Poussin (Figure 1c), that is not similar in terms of low level features but has the same subject and belongs to the same artistic movement.

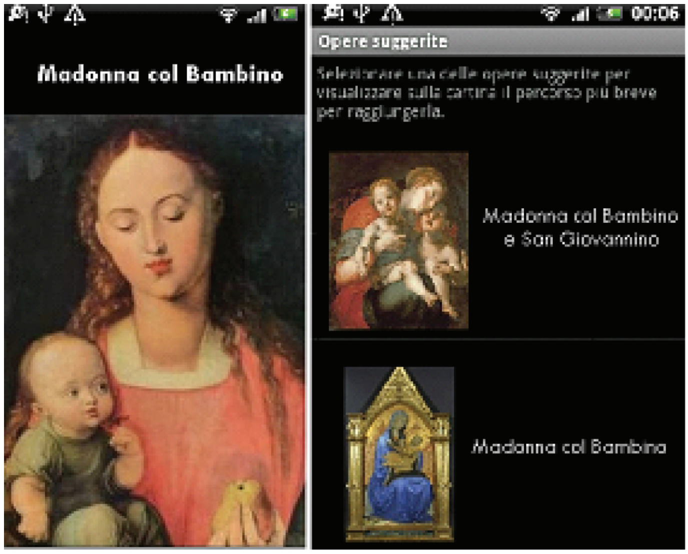
As another example, let us consider the case of a real museum (i.e., Uffizi Gallery in Florence) offering by means of a WiFi connection, a web-based access to a multimedia collection containing: digital reproductions of paintings, educational videos, audio guides, textual and hypermedia documents with description of authors and paintings. In order to make the user's experience in the museum more interesting and stimulating, the access to information should be customized based on the specific profile of a visitor, which includes learning needs, level of expertise and personal preferences, on user effective location in the museum, and on the "paintings similarity".

In particular, let us consider a user visiting the room 20 of the museum containing some paintings depicted by Albrecht Dűrer's and suppose that she is attracted, for example, by the painting entitled "Madonna col Bambino" (Figure 2). Also in this case, it would be helpful if the system could learn the preferences of the user (e.g., interests in paintings depicting the Holy Mary subject), based on the user current behaviour and past interactions of other museum visitors, and predict her future needs, by suggesting other paintings (or any other multimedia objects) representing the same or related subjects, depicted by the same or other related authors, or items that have been requested by users with similar preferences.

Similarly to the previous case, as an example, the user who is currently observing the Dűrer's painting in Figure 2 might be recommended to see in the room 27 a Jacopo Carucci's painting entitled "Madonna col Bambino e San Giovannino" (Figure 2), that is quite similar to the current picture in terms of color and texture, and in the room 3 "Madonna col Bambino" by Andrea Vanni (Figure 2), that is not very similar in terms of low level features but is more similar in terms of semantic content. Moreover, if in the past a lot of visitors after having seen the Dűrer's painting visited the room containing the special Pontormo and Rosso Fiorentino collection, the system could suggest visiting it and recommending some paintings within. If it is requested, the system could suggest by using museum maps and given the user location the way to reach each room from the current position.

During the visit, a user by her mobile device could read multimedia documents related to authors or paintings that she is viewing and

Figure 2. Paintings depicting the subject "Holy Mary"



listening to audio guides available for different languages.

Finally, the activities suggested to the user should be the following ones:

- *"See the painting Madonna col Bambino by Dûrer in the room 20"*
- *"See the painting Madonna col Bambino e San Giovannino by Carucci in the room 27"*
- *"See some paintings by Carrucci in the room 27"*

In both cases, from the *user perspective* there is the advantage of having a guide suggesting paintings which the users might be interested in, whereas, from the *system perspective*, there is the undoubted advantage of using the suggestions for pre-fetching and caching the objects that are more likely to be requested.

3. RELATED WORK AND BASIC IDEA BEHIND THE PROPOSED APPROACH

In the most common formulation, the *recommendation problem* is the problem of estimating *ratings* - sometimes called also *utilities* - for the set of items that has not yet been seen by a given user.

In *Content Based recommender systems* (Pazzani & Billsus, 2007), the utility r^i_j of item o_j is estimated using the utilities $r(u_i, o_k)$ assigned by the user u_i to items o_k that are in some way "similar" to item o_j. For example, in a movie recommendation application, in order to recommend movies to user *u*, the content-based recommender system tries to understand the commonalities among the movies user *u* has rated highly in the past (specific actors, directors, genres, subject matter, etc.). Then, only the movies that have a high degree of similarity to the user's preferences would be recommended.

One of the main drawbacks of these techniques is that they do not benefit from the great amount of information that could be derived by analyzing the behavior of other users. Moreover, the content must either be in a form that can be automatically parsed or the features should be assigned to items manually and, while information retrieval techniques work reasonably well in extracting features from text documents, some other domains, as multimedia, have an inherent problem with automatic feature extraction.

Another problem is that, if two different items are represented by the same set of features, they are indistinguishable. Eventually, a subtle problem is that the system can only recommend items that are similar to those already rated by the user itself (*overspecialization*).

Collaborative Filtering (Adomavicius & Tuzhilin, 2005) is, in the opposite, the process of filtering or evaluating items using the opinions of other people. Thus, unlike content-based recommendation methods, collaborative systems predict the utility of items r^i_j for a particular user u_i based on the utility $r(u_h,o_k)$ of items o_k previously rated by other users u_h "similar" to u_i. It takes its root from something human beings have been doing for centuries: sharing opinions with others. These opinions can be processed in real time to determine not only what a much larger community thinks of an item, but also develop a truly personalized view of that item using the opinions most appropriate for a given user or group of users (Resnick et al., 1994; Schafer et al., 2007; Kim et al., 2009).

The main problem behind collaborative filtering clearly is to associate each user to a set of other users having similar profiles. In order to make any recommendations, the system has to collect data mainly using two methods: the first one is to ask for explicit ratings from a user, while it is also possible to gather data implicitly logging actions performed by users. Once the data has been gathered, there are two basic ways of filtering through it to make predictions. The most basic method is *passive filtering*, which simply uses data aggregates to make predictions (such as the average rating for an item) and each user will be given the same predictions for a particular item (e.g., *digg.com*). In the opposite, *active filtering* uses patterns in user history to make predictions obtaining user-specific and context-aware recommendations (e.g., *Amazon*).

Collaborative systems have their own limitations that can be grouped under the name of the *cold start problem* that describes situations in which a recommender is unable to make meaningful recommendations due to an initial lack of ratings thus degrading the filtering performance. It can occur under three scenarios: *new user*, *new item,* and *new community*.

Content-based filtering and collaborative filtering may be manually combined by the end-user specifying particular features, essentially constraining recommendations to have certain content features. More often they are automatically combined in the so called *hybrid approach* (Basilico & Hofmann, 2004; Anand et al., 2007; Lam et al., 2008) that helps to avoid certain limitations of each method. Different ways to combine collaborative and content-based methods into a hybrid recommender system can be classified as follows: (i) implementing collaborative and content-based methods separately and combining their predictions; (ii) incorporating some content-based characteristics into a collaborative approach; (iii) incorporating some collaborative characteristics into a content-based approach; and (iv) constructing a general unifying model that incorporates both content-based and collaborative characteristics.

More recently, the discussed strategies have been extended to multimedia realm (e.g., multimedia repositories, digital libraries, multimedia sharing system, etc.) with the aim of considering in the more effective way multimedia content of recommended objects, both in terms of low-level and high-level characteristics (i.e., multimedia features and semantics), in the recommendation process together with user social behavior and preferences.

For what content-based techniques concerns, Maidel et al. (2008) proposes a method that exploits some ontologies for ranking items' relevancy in the electronic paper domain, while in Hijikata et al. (2006) a content based filtering has been applied to music data using decision trees. In the framework of multimedia sharing system, Musial et al. (2008) introduces a recommender system that use two ontologies (one for multimedia objects and one for users) in the context of a photo sharing system. To generate suggestions a new concept of "multirelational" social network was introduced, covering both direct as well as multimedia object-based relationships that reflect social and semantic links between users. Eventually, Manzato and Goularte (2009) propose a content-based recommender architecture which explores information that is available at the time users enhance content in order to capture a certain level of semantic information from the multimedia content and from user preferences that is at the base of their video recommender system.

Among collaborative-filtering proposals, authors in Baloian et al. (2004) propose a collaborative recommender system, which suggests multimedia learning material based on the learner's background and preferences. Kim et al. (2008) proposes a collaborative filtering-based multimedia contents recommender system in P2P architectures that rates multimedia objects of nearest peers with similar preference through peer-based local information only. Tseng et al. (2008) propose a system, which combines discovered relations between user preferences and conceptualized multimedia contents by annotation and association mining techniques, can provide a suitable recommendation list to assist users in making a decision among a massive amount of multimedia items (images, videos and music).

As hybrid solutions, the *uMender* system (Su & Ye, 2010) exploits context information, musical content and relevant user ratings to perform music recommendations on mobile devices. Knijnenburg et al. (2010) propose a user-centric approach to media recommendations that exploits subjective and objective measures of user experience, while users interact with multimedia data. A framework for recommendation of multimedia objects based on processing of individual ontologies is then proposed in Juszczyszyn et al. (2010): the recommendation process takes into account similarities calculated both between objects (metadata) and users' ontologies, which reflect the social and semantic features existing in the system. Another example of hybrid approach, implemented in *MoRe*, a movie recommendation system, has been described in Lekakos and Caravelas (2008). Finally, low and high level features have been used to define the similarity among multimedia items in Albanese, Chianese et al. (2010); this measure is then used to compare patterns of past users in order to identify users with similar browsing behavior.

As we can note, the majority of approaches to recommendation in the multimedia realm exploits high level metadata - extracted in automatic or semi-automatic way from low level features - that are in different manners correlated and compared with user preferences, usually mapped in the shape of ontologies. These approaches suffer from some drawbacks:

- It is not always possible to extract in automatic and effective way useful high level information from multimedia features (automatic annotation algorithms have not always high performances);
- For some kinds of multimedia data there is not a precise correlation between high and low level information (e.g., in images the concept of "moon" is related to a region with a circular shape and white color with a given *uncertainty*);
- There are not always available explicit and useful information (knowledge) about user preferences (e.g., usually a user can retrieve information from a multimedia system without the necessity of a registration, as in youtube or flickr);
- In the recommendation process and for particular kinds of multimedia data sometimes it is useful to take into account features of the objects that user is currently observing as content information (e.g., the main colours of a painting are often an indication of the related artistic movement or school).

Our approach tries to avoid such drawbacks:

- Supposing the existence of a-priori knowledge about metadata values and their relationships (i.e., a taxonomy is used to define high-level concepts);
- Considering in a separate way low and high level information, i.e., both contribute to determine the utility of an object in the recommendation process;

- Exploiting system logs to implicitly determine information about a user and the related community and considering their browsing sessions as a sort of ratings;
- Considering as relevant content for the recommendation the features of the object that a user is currently watching.

Thus, we try to meet in a unique strategy some aspects of multimedia information retrieval with the basic theory of modern recommender systems.

4. THE RECOMMENDATION STRATEGY

An effective multimedia recommender system for supporting intelligent browsing of multimedia collections has the capability of reliably identifying the objects that are most likely to satisfy the interests of a user at any given point of her exploration. In our case, we have to address four fundamental questions:

1. How can we select a set of objects from the collection that are good candidates for recommendation?
2. How can we rank the set of candidates?
3. How can we capture, represent and manage semantics related to multimedia objects to reduce the semantic gap between what the user is watching and what she is looking for?
4. How can we take into account such semantics in the recommendation process?

To give an answer to the first two questions, we have based our recommendation algorithm on an importance ranking method that strongly resembles the *PageRank* ranking system (Albanese, d'Acierno et al., 2010) and model recommendation as a social choice problem, proposing a method that computes customized recommendations by originally combining several features of multimedia objects (low-level and semantics), past behavior of individual users and overall behavior of the entire community of users. Our basic idea is to assume that when an object o_i is chosen after an object o_j in the same browsing session, this event means that o_j "is voting" for o_i. Similarly, the fact that an object o_i is very similar to o_j can also be interpreted as o_j "recommending" o_i (and viceversa).

Thus, our idea is to model a browsing system for a set of object O as a labeled graph (G,l), where $G=(O,E)$ is a directed graph and $l: E \rightarrow \{pattern, sim\} \times R^+$ is a function that associates each edge in $E \subseteq O \times O$ with a pair (t,w), where t is the type of the edge which can assume two enumerative values (*pattern* and *similarity*) and w is the weight of the edge. According to this model, we can list two different cases:

- A *pattern label* for an edge (o_j,o_i) denotes the fact that an object o_i was accessed immediately after an object o_j and, in this case, the weight w^i_j is the number of times o_i was accessed immediately after o_j .
- The *similarity label* for an edge (o_j,o_i) denotes the fact that an object o_i is similar to o_j and, in this case, the weight w^i_j is the similarity between o_j and o_i.

Thus, a link from o_j to o_i indicates that part of the importance of o_j is transferred to o_i. Given a labeled graph (G,l), we can formulate the definition of recommendation grade of a multimedia object more formally as follows.

Definition 3.1: (Recommendation Grade ρ)

$$\forall o_i \in O \ \rho(o_i) = \sum_{o_j \in PG(o_i)} w^i_j \, o_j \rho(o_j) \qquad (1)$$

where $P_G=\{o_j \in O | (o_j, o_i) \in E\}$ is the set of predecessors of o_i in G, and w^i_j is the normalized weight of the edge from o_j to o_i. For each $o_j \in O$, $\sum_{o_i \in SG(o_j)} w^i_j = 1$ must hold, where $S_G=\{o_i \in O | (o_j,o_i) \in E\}$ is the set of successors of o_j in G.

It is easy to see that the vector $R = [\rho(o_i) \ldots \rho(o_n)]^T$ can be computed as the solution to the following equation:

$$R = C \cdot R \tag{2}$$

where $C=\{w^i_j\}$ is an ad-hoc matrix that defines how the importance of each object is transferred to other objects and can be seen as a linear combination of the following elements (Albanese et al., 2011):

- A *local browsing matrix* $A_l=\{a^l_{ij}\}$ for each user $u_l \in U$. Its generic element a^l_{ij} is defined as the ratio of the number of times object o_i has been accessed by user u_l immediately after o_j to the number of times any object in O has been accessed by u_l immediately after o_j.
- A *global browsing matrix* $A=\{a_{ij}\}$. Its generic element a_{ij} is defined as the ratio of the number of times object o_i has been accessed by any user immediately after o_j to the number of times any object in O has been accessed immediately after o_j.
- A *multimedia similarity matrix* $B=\{b_{ij}\}$ such that $b_{ij}=\sigma(o_i,o_j)/\Gamma$ if $\sigma(o_i,o_j)\geq\tau \ \forall i\neq j$, 0 otherwise. σ is any similarity function defined over O which calculates for each couple of objects their multimedia relatedness in terms of low (features) and high level (semantics) descriptors; τ is a threshold and Γ is a normalization factor which guarantees that $\sum_i b_{ij}=1$.

The introduction of matrix *B* allows to address the two last questions that we introduced at the beginning of the section and thus to introduce a sort of content-based image retrieval with high-level semantics in the recommendation process. In particular, to compute B matrix, we have decided to adopt 5 sets of the most diffused multimedia features (Tamura descriptors, MPEG-7 color-based descriptors, MPEG-7 edge-based descriptors, MPEG-7 color layout-based descriptors and all MPEG7 descriptors (Lux & Chatzichristofis, 2008) and the related similarity metrics have been implemented by LIRE tool. In addition, we exploit specific image metadata (*artist*, *genre* and *subject*) and the semantic similarity has been computed used the most diffused metrics for semantic relatedness of concepts based on a vocabulary (Li-Bandar-McLean, Wu-Palmer, Rada, Leacock-Chodorow) (Budanitsky & Hirst, 2001). In particular the semantic similarity combines similarities among artists, genres and subjects obtained by using a fixed taxonomy produced by domain experts with image features.

In Albanese et al. (2011) the experimental protocol to determine the best combination of the proposed metrics is reported for images representing artistic paintings. In particular, the combination between high and low level descriptors is Sugeno fuzzy integral of Li and MPEG-7 color layout-based similarities in order to have more high values of precision, and Sugeno fuzzy integral of Wu-Palmer and MPEG-7 color based similarities in order to have more high level values of recall, thus we use this combination for matrix *B* computation.

So far we have a suitable manner to represent object features and to compare the related similarity also considering semantics in terms of object metadata; now, our main goal is to compute customized rankings for each individual user.

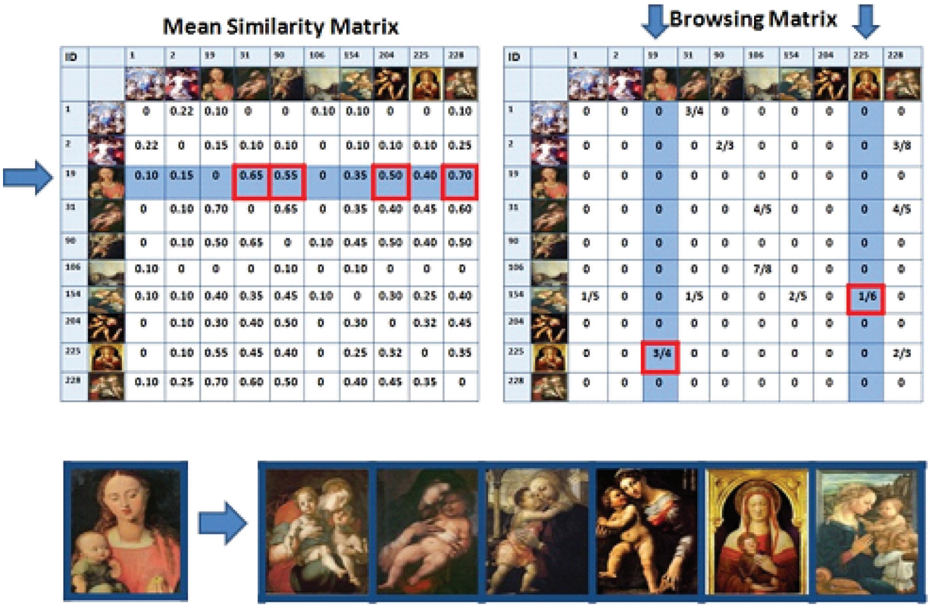
In this case, we can then rewrite equation 2 considering the ranking for each user as follows:

$$R_l = C \cdot R_l \tag{3}$$

where $R_l=[\rho(o_i)\ldots\rho(o_n)]^T$ is the vector of recommendation grades, customized for a user u_l.

We note that solving equation 3 corresponds to finding the stationary vector of *C*, i.e., the eigenvector with eigenvalue $\lambda = 1$. We demonstrated in Albanese, d'Acierno et al. (2010) - here all the details for computation of recommendation grades are also reported - that *C*, under certain assumptions and transformations, is a real square matrix having positive elements, with a unique largest real eigenvalue and the corresponding eigenvector has strictly positive

Figure 3. How to compute the set of candidates



Mean Similarity Matrix

ID	1	2	19	31	90	106	154	204	225	228
1	0	0.22	0.10	0	0	0.10	0.10	0	0	0.10
2	0.22	0	0.15	0.10	0.10	0	0.10	0.10	0.10	0.25
19	0.10	0.15	0	0.65	0.55	0	0.35	0.50	0.40	0.70
31	0	0.10	0.70	0	0.65	0	0.35	0.40	0.45	0.60
90	0	0.10	0.50	0.65	0	0.10	0.45	0.50	0.40	0.50
106	0.10	0	0	0	0.10	0	0.10	0	0	0
154	0.10	0.10	0.40	0.35	0.45	0.10	0	0.30	0.25	0.40
204	0	0.10	0.30	0.40	0.50	0	0.30	0	0.32	0.45
225	0	0.10	0.55	0.45	0.40	0	0.25	0.32	0	0.35
228	0.10	0.25	0.70	0.60	0.50	0	0.40	0.45	0.35	0

Browsing Matrix

ID	1	2	19	31	90	106	154	204	225	228
1	0	0	0	3/4	0	0	0	0	0	0
2	0	0	0	0	2/3	0	0	0	0	3/8
19	0	0	0	0	0	0	0	0	0	0
31	0	0	0	0	0	4/5	0	0	0	4/5
90	0	0	0	0	0	0	0	0	0	0
106	0	0	0	0	0	7/8	0	0	0	0
154	1/5	0	0	1/5	0	0	2/5	0	1/6	0
204	0	0	0	0	0	0	0	0	0	0
225	0	0	3/4	0	0	0	0	0	0	2/5
228	0	0	0	0	0	0	0	0	0	0

components. In such conditions, equation 3 can be solved used the *Power Method* algorithm.

The matrix C does not have to be computed for all database objects, but only for those objects that are good candidates to recommendation. Assuming that a user u_l is currently watching object o_j, we can generate the final set of candidate recommendations by considering:

$$C = \cup_{k=1 \ldots M} (\{o_i \in O | A^k_{ij} > 0\} \cup \{ o_{i \in} O | B_{ij} > 0\}) \quad (4)$$

The set of candidates includes the objects that have been accessed by at least one user within k steps from o_j, with k between 1 and M, and the objects that are most similar to o_j. Note that a positive element a^k_{ij} of A^k indicates that o_i was accessed exactly k steps after o_j at least once.

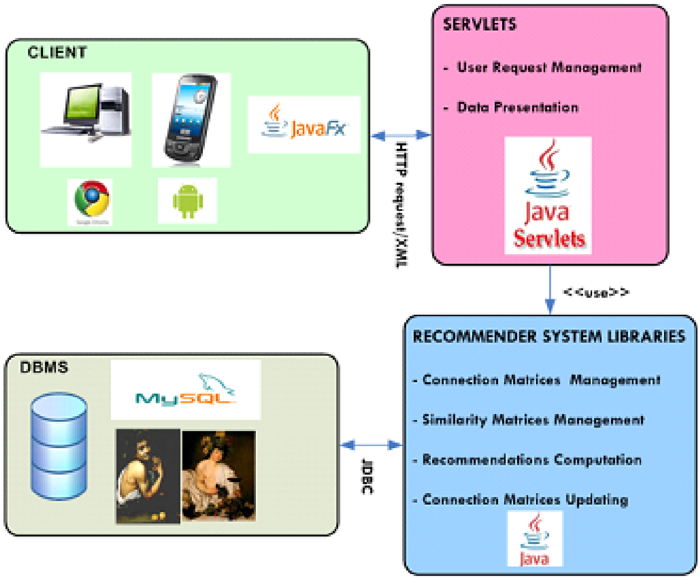
In Figure 3 there is an easy example of how to compute the set of candidates in the case the collection has only ten paintings and the most similar images to the current image are only four. As we can see, after selecting the most similar images, from each of these the images accessed within 2 steps are selected.

5. THE SYSTEM

In this section, we describe a case study in the cultural heritage domain for a web recommendation system that provides browsing facilities for a multimedia collection of paintings. In particular, our recommender helps the users for finding paintings of interest from a large set of choices, proposing a set of suggestions for each observed object; the recommendations are computed combining the user's behaviour with low and high level image descriptors, following the previous described approach.

We use a *memory-based* algorithm so that low and high level similarities are evaluated once; this reflect the unchanging nature of these measures while, clearly, if we add new paintings, similarity matrices have to be conveniently updated. Instead, to capture the dynamic nature of user's behaviour, we periodically re-compute connection matrices;

Figure 4. The system architecture



specifically, each connection matrix is updated as soon the browsing session ends. To solve the cold start problem, when there is no information about user's behaviour, our system uses low or/and high level similarities, in addition to the extracted behaviour of the whole community. For new items, of course, recommendation is based just on similarities.

Our data collection can consist of different digital reproductions of paintings (managed by *MySQL* DBMS) to which it is possible to associate artists and artistic genres. Each painting can be also linked to a pair of subjects, chosen among a list containing the available ones; such an information roughly represents what the painting represents. A user can interact with our system (Figure 4) using a web browser that communicates with the server by means of straight http requests or by means of an *Android* application. In the first case, the presentation logic is based on *JavaFX* technology that allows interaction with users using advanced graphical functionalities. The client requests are elaborated by *JAVA Servlets* and results are sent to the client in form of XML data (according to the *Service Oriented Architecture* paradigm).

The core functionality of the system, the recommendation process, can be described as follow.

As soon as a user interacts with the system, the core process starts defining the set of candidates for the recommendation by considering the union of:

- The set of paintings which are the most similar to the current one, according to the similarity matrices;
- The set of paintings which have been accessed by at least one user within a certain number of steps from the current one; to reach this goal, if the user is logged in and there exits the related user connection matrix, the past user's behaviour is considered; otherwise the global connection matrix is taken into account.

The set of candidates, of course, takes into account the user's context and, thus, the *C* matrix is built just referencing the elements belonging to such set; the Power Method is then used to compute the ranking vector, that is in eventually exploited to recommend new paintings. At the end of each browsing session, the system updates the connection matrix extracting a set of pairs (image to be accessed - accessed image); for instance, if once observed the j-th painting the user sees the i-th painting, then an occurrence will be added to the A_{ij} element in the corresponding connection matrix.

From the final users perspective, the client application has the following features:

- A set of forms to provide users log in or registration;
- A gallery to visualize images which are returned after a search by author, subject or artistic genre;
- Visualization of an image and of the related information and multimedia presentation of recommended images;
- Storing of user session with the information related to the browsing patterns.

The client side of our system can be adapted both for a mobile and PC device and customized for different applications, while the server side can host different multimedia collection of images (e.g., *Uffizi* and *Olga's* galleries). Some screenshots related to the client application realized for browsing of a real museum are reported in Figure 5.

6. PRELIMINARY EXPERIMENTAL RESULTS

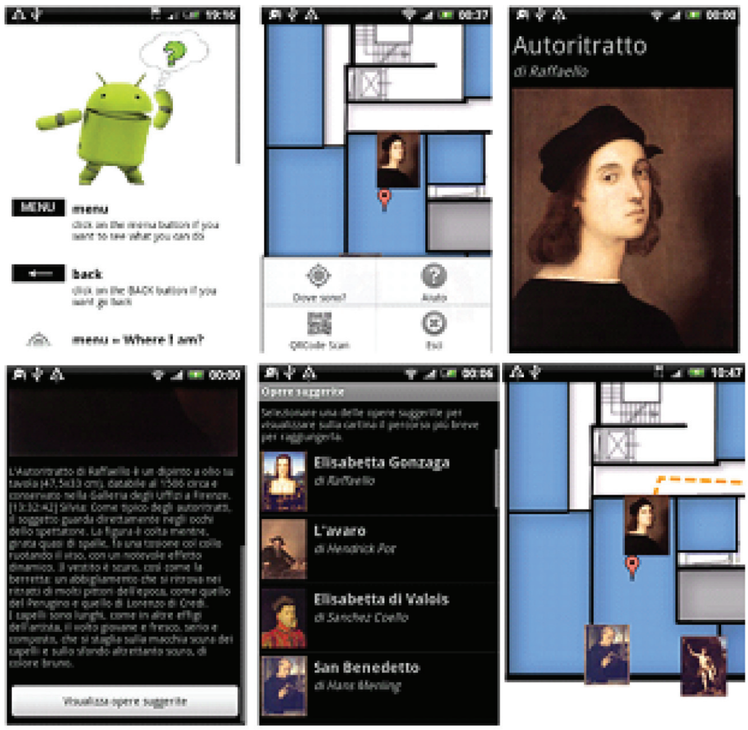
Recommender systems are complex applications that are based on a combination of several models, algorithms and heuristics. This complexity makes evaluation efforts very difficult and results are hardly generalizable, which is apparent in the literature about recommender evaluation (Schulz & Hahsler, 2002). Previous research work on recommender system evaluation has mainly focused on algorithm *accuracy*, especially objective prediction accuracy. More recently, researchers began examining issues related to users' subjective opinions and developing additional criteria to evaluate recommender systems. In particular, they suggest that user's satisfaction does not always (or, at least, not only) correlate with the overall recommender's accuracy and evaluation frameworks for measuring the perceived qualities of a recommender and for predicting user's behavioural intentions as a result of these qualities should be taken into account.

In Albanese et al. (2011) a user-centric evaluation is proposed and reported some preliminary experimental results about user satisfaction in using a recommendation strategy for browsing the *Uffizi Gallery* (containing only 474 paintings). The goal was to establish how helpful our system was to provide an exploration of digital reproductions of paintings. Moreover from these experiments we wanted to understand how helpful recommendations offered by our recommender system were to address users toward paintings which satisfied their interests. In particular, it has been demonstrated that the introduction of recommendation techniques can improve the system usability with respect to assigned browsing tasks and we evaluated such an improvement in terms of empirical measurements of access complexity and *TLX* factors (with regards to a system that does not exploit recommendation, i.e., Picasa) provided by different kinds of users.

In this paper we want to repeat the proposed experimental protocol, adding an evaluation on browsing effectiveness, using a more large multimedia collection, the paintings belonging to the *Olga's Gallery* on-line art museum; in particular, the collection includes about 10,000 paintings encompassing 28 genres (e.g., Cubism, Baroque, Early Renaissance), more than 200 authors (e.g., Caravaggio, Rubens), and about 100 subjects (e.g., Landscapes, Portraits).

In the training phase, we have chosen 20 users among students and graduate students that used for 5 days the system without recommendation facilities to capture their browsing sessions and build a consistent matrix A. As comparing system we consider the web site of Olga's Gallery (http://www.abcgallery.com) that provides a classical keywords-based search engine and some indexes (on artist and genre

Figure 5. Screenshots of the client application



names) to facilitate retrieval of painting of interest. Finally we have set $k=3$ for computing the candidates set and used a (sliding) window of 50 recommended items.

Browsing Effectiveness

This first set of experiments aims at comparing the ranking provided by our system using the proposed recommendation degree with the ranking provided by a human observer. To this end, we have slightly modified a test proposed by Santini (2000), in order to evaluate the difference between the two rankings ("treatments") in terms of hypothesis verification on the entire dataset.

Consider a weighted displacement measure defined as follows.

Let q be a query (example painting) on a database of n images that produces k results (recommended items). There is one ordering (usually given by one or more human subjects) which is considered as the ground truth, represented as:

$$R_h = \{o_1, \ldots, o_k\} \tag{5}$$

Every image in the ordering has also associated a measure of relevance $0 \leq S(o, q) \leq 1$ such that (for the ground truth), $S(o_i, q) \geq S(o_{i+1}, Q)$, $\forall i$. This is compared with an (experimental) ordering:

$$Rs = \{o_{\pi_1}, \ldots, o_{\pi_k}\} \tag{6}$$

where $\{\pi_1, \ldots, \pi_k\}$ is a permutation of $1, \ldots, n$. The displacement of o_i is defined as:

$$d_q(o_i) = |i - \pi_i| \quad (7)$$

While, the relative weighted displacement of *Rs* is defined as:

$$W_q = \sum_i (S(o_i, q) \cdot d_q(o_i))/\Omega \quad (8)$$

$\Omega = \lceil n^2/2 \rceil$ being a normalization factor.

Relevance *S* is obtained from the subjects asking them to divide the results in three groups: *very similar* ($S(o_i, q) = 1$), *quite similar* ($S(o_i, q) = 0.5$) and *dissimilar* ($S(o_i, q) = 0.05$).

In our experiments, on the basis of the ground truth provided by human subjects, treatments provided either by humans or by our system are compared. The goal is to determine whether the observed differences can indeed be ascribed to the different treatments or are caused by random variations. In terms of hypothesis verification, if μ_i is the average score obtained with the *i-th* treatment, a test is performed in order to accept or reject the null hypothesis H_0 that all the averages μ_i are the same (i.e., the differences are due only to random variations).

Clearly the alternate hypothesis H_1 is that the means are not equal, that is the experiment actually revealed a difference among treatments. The acceptance of the H_0 hypothesis can be checked with the *F* ratio. Let us assume that there are *M* treatments and *N* measurements (experiments) for each treatment. Let us define:

- t_{ij} the result of the *j*-th experiment performed with the *i*-th treatment in place;
- $\mu_i = (\sum_{j=1,N} t_{ij})/N$ the average for treatment t_{ij};
- $\mu = (\sum_{i=1,M} \mu_i)/M$ the total average;
- $\sigma^2_A = (\sum_{i=1,M} (\mu_i - \mu)^2)/(N/(M-1))$ the between treatments variance;
- $\sigma^2_W = ((\sum_{i=1,M} \sum_{j=1,N} (t_{ij} - \mu_i)^2)/(M/(N-1))$ the within treatments variance.

Then, the *F* ratio is:

$$F = \sigma^2_A / \sigma^2_W \quad (9)$$

A high value of *F* means that the between treatments variance is preponderant with respect to the within treatment variance, that is, that the differences in the averages are likely to be due to the treatments.

In our experiments we employed 12 subjects, experts on art. Ten users, randomly chosen among the 12, were employed to determine the ground truth ranking and the other two served to provide the treatments to be compared with our system. 10 example images were selected, and for each of them we run the recommendation algorithm returning a result set of ten best candidates, for a total of 100 objects.

Result sets were randomly ordered to prevent bias and the two students were then asked to rank images in each set in terms their level of recommendation with respect to the query object. Each subject was also asked to divide the ranked objects in three groups: the first group consisted of images judged *very relevant* to the query, the second group consisted of images judged *quite relevant* to the query, and the third of *non relevant* images.

The mean and variance of the weighted displacement of the two subjects and of our system with respect to the ground truth are reported in Table 1. Then, the *F* ratio for each pair of distances was computed in order to establish which differences were significant. As can be noted from Table 2, the *F* ratio is always less than 1 and since the critical value F_0, regardless of the confidence degree (the probability of rejecting the right hypothesis), is greater than 1, the null hypothesis can be statistically accepted.

User Satisfaction

In order to evaluate the impact of the system on the users we have conducted the following experiments. We asked a different group of about 20 people (all medium experts in art) to browse a collection of images and complete several browsing tasks (20 tasks per user) of different complexity (five tasks for each complexity level), using the Olga's Gallery web site. After this test, we asked them to browse once again

Table 1. Mean and variance of the weighted displacement for the three treatments (two human subjects and system)

	Human 1	Human 2	System
μ_i	0.0475	0.0412	0.0304
σ^2_i	$8.266e^{-4}$	$8.895e^{-4}$	$8.951e^{-4}$

Table 2. The F ratio measured for pairs of distances (human vs. human and human vs. system)

F	Human 1	Human 2	System
System	0.484	0.713	0
Human 2	0.0899	0	
Human 1	0		

the same collection with the assistance of our recommender system and complete other 20 tasks of the same complexity.

We have subdivided browsing tasks in the following four broad categories:

- *Low Complexity* tasks (*Q*1)—e.g., "explore at least 10 paintings of *Baroque* style authored by *Caravaggio* and depicting a *religious* subject";
- *Medium Complexity* tasks (*Q*2)—e.g., "explore at least 20 paintings of *Baroque* authors that have *nature* as their subject";
- *High Complexity* tasks (*Q*3)—e.g., "explore at least 30 paintings of *Baroque* authors with subject *nature* and with a predominance of *red color*";
- *Very High Complexity* tasks (*Q*4)—e.g., "explore at least 50 paintings of *Baroque* authors depicting a *religious* subject with a predominance of *red color*".

Note that the complexity of a task depends on several factors: the number of objects to explore, the type of desired features (either low or high-level), and the number of constraints (genre, author, subject). Two strategies were used to evaluate the results of this experiment: empirical measurements of access complexity in terms of *mouse clicks* and *time*, and *TLX* (*NASA Task Load Index* factors).

With respect to the first strategy, we measured the following parameters:

- *Access Time (t_a)*. The average time spent by the users to request and access all the objects for a given class of tasks.
- *Number of Clicks (n_c)*. The average number of clicks necessary to collect all the requested objects for a given class of tasks.

Table 3 reports the average values of t_a and n_c, for both Olga's Gallery and our system, for each of the four task complexity levels defined earlier.

In the second experiment, we asked the users to express their opinion about the capability of Olga's gallery and our system respectively to provide an effective user experience in completing the assigned browsing tasks. To this end, we used the TLX evaluation form to assess the workload on operators of various human–machine systems. Specifically, TLX is a multi-dimensional rating procedure that provides an overall workload score based on a weighted average of ratings on six sub-scales: mental demand, physical demand, temporal demand, own performance, effort and frustration (lower TLX scores are better). In other words, this experiment was aimed at measuring how difficult is for a user to use either our system or Olga's gallery to complete a browsing task. We obtained the average result scores

Table 3. Comparison between our system and Olga's gallery in terms of t_a and n_c

Task Class	System	t_a (sec.)	n_c
Q1	Olga	58.7	14.5
Q1	Our System	53.8	13.2
Q2	Olga	174.2	41.8
Q2	Our System	62.5	21.5
Q3	Olga	309.2	77.1
Q3	Our System	145.1	38.2
Q4	Olga	500.7	144.2
Q4	Our System	232.9	56.7

Table 4. Comparison between our system and Olga's gallery in terms of TLX factors

TLX Factor	Our System	Olga's Gallery
Effort	49.3	60.5
Mental Demand	53.2	58.3
Physical Demand	44.4	50.2
Temporal Demand	50.6	70.1
Frustration	58.3	72.8
Own Performance	39.5	44.2

reported in Table 4, which show that our system outperforms Olga in every sub-scale.

It is evident that the two aspects where our system beats Olga by the largest margin are temporal demand and frustration. This result implies that our system completes browsing tasks faster and provides a better (less frustrating) user experience. In addition, the fact that browsing tasks can be completed faster using our system is an indication that recommendations are effective, as they allow a user to explore interesting and related objects one after another, without the interference of undesired items that would necessarily slow down the process.

7. CONCLUSIONS AND FUTURE WORK

In this paper we proposed a novel approach to recommendation for multimedia browsing systems, based on a method that computes customized recommendations by combing in an original way intrinsic features (semantic contents and low-level features) of the objects, past behaviour of individual users and behaviour of the users' community as a whole. In particular, we realized a recommender system which helps users to browse digital reproductions of Uffizi paintings, providing them suggestions computed by our novel method for recommendations. Then we investigated the effectiveness of the proposed approach in the considered scenario, based on the browsing effectiveness and user satisfaction. Experimental results showed that our approach is promising and encourages further research in this direction.

Future works will be devoted to: (i) introduce explicit user profiling mechanism based on the creation of users' categories, (ii) extend experimentation on a larger image data set, and

(iii) compare our algorithm with respect to other recommendation strategies.

REFERENCES

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, *17*(6), 734–749. doi:10.1109/TKDE.2005.99

Albanese, M., Chianese, A., d'Acierno, A., Moscato, V., & Picariello, A. (2010). A multimedia recommender integrating object features and user behavior. *Multimedia Tools and Applications*, *50*(3), 563–585. doi:10.1007/s11042-010-0480-8

Albanese, M., d'Acierno, A., Moscato, V., Persia, F., & Picariello, A. (2010). Modeling recommendation as a social choice problem. In *Proceedings of the 4th ACM Conference on Recommender Systems* (pp. 329-332).

Albanese, M., d'Acierno, A., Moscato, V., Persia, F., & Picariello, A. (2011). A multimedia semantic recommender system for cultural heritage applications. In *Proceedings of the 5th IEEE Conference on Semantic Computing – Semantic Multimedia Management Workshop.*

Anand, S. S., Kearney, P., & Shapcott, M. (2007). Generating semantically enriched user profiles for web personalization. *ACM Transactions on Internet Technology*, *7*(4). doi:10.1145/1278366.1278371

Baloian, N. A., Galdames, P., Collazos, C. A., & Guerrero, L. A. (2004). A model for a collaborative recommender system for multimedia learning material. In G.-J. de Vreede, L. A. Guerrero, & G. M. Raventos (Eds.), *Proceedings of the 10th International Workshop on Groupware: Design, Implementation, and Use* (LNCS 3198, pp. 281-288).

Basilico, J., & Hofmann, T. (2004). Unifying collaborative and content-based filtering. In *Proceedings of the 21st Conference on Machine Learning*.

Budanitsky, A., & Hirst, G. (2001). Semantic distance in Wordnet: An experimental, application oriented evaluation of five measures. In *Proceedings of the Workshop on WordNet and Other Lexical Resources*.

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, *22*(1), 5–53. doi:10.1145/963770.963772

Hijikata, Y., Iwahama, K., & Nishida, I. S. (2006). Content-based music filtering system with editable user profile. In *Proceedings of the ACM Symposium on Applied Computing* (pp. 1050-1057).

Juszczyszyn, K., Kazienko, P., & Musiał, K. (2010). Personalized ontology-based recommender systems for multimedia objects, agent and multi-agent technology for internet and enterprise systems. In Hakansson, A., Hartung, R., & Nguyen, N.-T. (Eds.), *Agent and multiagent technology for Internet and enterprise systems* (Vol. 289, pp. 275–292). Berlin, Germany: Springer-Verlag. doi:10.1007/978-3-642-13526-2_13

Kazienko, P., & Musial, K. (2006). Recommendation framework for online social networks. *Advances in Web Intelligence and Data Mining*, *23*, 111–120. doi:10.1007/3-540-33880-2_12

Kim, H. K., Kim, J. K., & Ryu, Y. U. (2009). Personalized recommendation over a customer network for ubiquitous shopping. *IEEE Transactions on Services Computing*, *2*(2), 140–151. doi:10.1109/TSC.2009.7

Kim, J. K., Kim, H. K., & Cho, Y. H. (2008). A user-oriented contents recommendation system in peer to peer architecture. *Expert Systems with Applications*, *34*(1), 300–312. doi:10.1016/j.eswa.2006.09.034

Knijnenburg, B., Meesters, L., Marrow, P., & Bouwhuis, D. (2010). User-centric evaluation framework for multimedia recommender systems. In P. Daras & O. M. Ibarra (Eds.), *Proceedings of the First International Conference on User Centric Media* (LNICST 40, pp. 366-369).

Lam, X. N., Vu, T., Le, T. D., & Duong, A. D. (2008). Addressing cold-start problem in recommendation systems. In *Proceedings of the 2nd International ACM Conference on Ubiquitous Information Management and Communication* (pp. 208-211).

Lekakos, G., & Caravelas, P. (2008). A hybrid approach for movie recommendation. *Multimedia Tools and Applications*, *36*(1-2), 55–70. doi:10.1007/s11042-006-0082-7

Lux, M., & Chatizichristofis, A. (2008). LIRE: Lucene Image REtrieval - an extensible java cbir library. In *Proceedings of the 16th ACM International Conference on Multimedia* (pp. 1085-1088).

Maidel, V., Shoval, P., Shapira, B., & Taieb-Maimon, M. (2008). Evaluation of an ontology-content based filtering method for a personalized newspaper. In *Proceedings of the ACM Conference on Recommender Systems* (pp. 91-98).

Manzato, M. G., & Goularte, R. (2009). Supporting multimedia recommender systems with peer-level annotations. In *Proceedings of the 15th Brazilian Symposium on Multimedia and the Web*.

Musiał, K., Kazienko, P., & Kajdanowicz, T. (2008). Social recommendations within the multimedia sharing systems. In M. D. Lytras, J. M. Carroll, E. Damiani, & R. D. Tennyson (Eds.), *Proceedings of the First World Summit on Emerging Technologies and Information Systems for the Knowledge Society* (LNCS 5288, pp. 364-372)

Pazzani, M., & Billsus, D. (2007). Content-based recommendation systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *Proceedings of the Methods and Strategies of Web Personalisation: The Adaptive Web* (LNCS 4321, pp. 325-341).

Qin, S., Menezes, R., & Silaghi, M. (2010). A recommender system for Youtube based on its network of reviewers. In *Proceedings of the IEEE Second International Conference on Social Computing* (pp. 323-328).

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: an open architecture for collaborative filtering of netnews. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (pp. 175-186).

Santini, S. (2000). Evaluation vademecum for visual information systems. In *Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases* (Vol. 3972, pp. 132-143).

Schafer, J. B., Frankowsky, D., Herlocker, J. L., & Sen, S. (2007). Collaborative filtering recommender systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *Proceedings of the Methods and Strategies of Web Personalisation: The Adaptive Web* (LNCS 4321, pp. 291-324).

Schulz, A. G., & Hahsler, M. (2002). Evaluation of recommender algorithms for an internet information broker based on simple association rules and on the repeat-buying theory. In *Proceedings of the Fourth WebKDD Workshop: Web Mining for Usage Patterns & User Profiles* (pp. 100-114).

Su, J. W., & Yeh, H. H. (2010). Music recommendation using content and context information mining. *IEEE Intelligent Systems*, *25*(1), 16–26. doi:10.1109/MIS.2010.23

Tseng, V. S., Su, J.-H., Wang, B.-W., Hsiao, C.-Y., Huang, J., & Yeh, H.-H. (2008). Intelligent multimedia recommender by integrating annotation and association mining. In *Proceedings of the IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing* (pp. 492-499).

Zhang, X., & Wang, H. (2005). Study on recommender systems for business-to-business electronic commerce. *Communications of the IIMA*, *5*(4), 53–61.

Massimiliano Albanese received a Laurea degree in computer science and engineering from the University of Naples "Federico II" in 2002. In 2005, he received his PhD degree in Computer Science and Engineering from the same University, where he then served until 2006 as a research and teaching assistant with the Multimedia Information Systems Group. He joined in 2006 the University of Maryland Institute for Advanced Computer Studies, College Park, as a post doctoral researcher and, in 2011, he joined the Center of Secure Info Systems of George Mason University as an Assistant Professor. His primary areas of interest are in activity detection, computer security, multimedia databases, and knowledge representation and management.

Antonio d'Acierno received the Laurea degree com laude in electronics engineering from the University of Naples Federico II . Since 1988 to 1999, he was actively integrated into the research group of IRSIP (Institute for Research on Parallel Informatics Systems) of the National Research Council of Italy (CNR). In 1999 he joined the Institute of Food Science (ISA) of CNR. He is currently a contract professor of Data Base Systems at the Faculty of Computer Science Engineering of the University of Salerno. His current research interests lie in the field of mobile transactions, information retrieval, semantic web, multimedia ontologies and applications, multimedia knowledge extraction and management, and bioinformatics.

Vincenzo Moscato received the Laurea degree (cum laude) in computer science and engineering from the University of Naples "Federico II", Italy, in 2002. In 2005, he received the PhD degree in computer science and engineering at the same University. In 2009 he joined the Dipartimento di Informatica e Sistemistica of University of Napoli "Federico II", where he is currently an assistant professor of data base and computer engineering. He has been active in the field of computer vision, video and image indexing and multimedia data sources integration. His current research interests lie in the area of multimedia databases, video-surveillance applications and knowledge representation and management.

Fabio Persia received in 2009 the Laurea degree in computer science and engineering from the University of Naples "Federico II", Italy, where he is currently a PhD student in computer science and engineering. His current research interests lie in the field of video surveillance applications, multimedia databases and knowledge representation and management.

Antonio Picariello received the laurea degree in electronics engineering and the PhD degree in computer science and engineering from the University of Naples "Federico II," Italy, in 1991 and 1998, respectively. In 1993, he joined the Istituto Ricerca sui Sistemi Informatici Paralleli, National Research Council, Naples, Italy. In 1999, he joined the Dipartimento di Informatica e Sistemistica, University of Naples "Federico II" and is currently an associate professor of databases and computer engineering. He has been active in the field of computer vision, medical image processing and pattern recognition, object-oriented models for image processing, multimedia databases, and information retrieval. His current research interests include knowledge extraction and management, multimedia integration, and image and video databases.