Mining Networked Media Collections

Stephane Marchand-Maillet, Donn Morrison, Eniko Szekely,
Jana Kludas, Marc von Wyl and Eric Bruno*

Viper group – University of Geneva, Switzerland
http://viper.unige.ch
Stephane.Marchand-Maillet@unige.ch

Abstract. Multimedia data collections immersed into social networks may be explored from the point of view of varying documents and users characteristics. In this paper, we develop a unified model to embed documents, concepts and users into coherent structures from which to extract optimal subsets and to diffuse information. The result is the definition information propagation strategies and of active guiding navigation strategies of both the user and document networks, as a complement to classical search operations. Example benefits brought by our model are provided via experimental results.

1 Introduction

Many current information management systems are centered on the notion of a query related to information search. This is true over the Web (with all classical Web Search Engines), and for Digital Libraries. In the domain of multimedia, available commercial applications propose rather simple management services whereas research prototypes are also looking at responding to queries. In the most general case, information browsing is designed to supplement search operations. This comes from the fact that the multimedia querying systems largely demonstrate their capabilities using query-based scenario (by Example, by concepts) and these strategies often show limitations, be it in their scalability, their usability or utility or their capabilities or precision. Multimedia search systems are mostly based on content similarity. Hence, to fulfill an information need, the user must express it with respect to relevant (positive) and non-relevant (negative) examples. From there, some form of learning is performed, in order to retrieve the documents that are the most similar to the combination of relevant examples and dissimilar to the combination of non-relevant examples. The question then arises of how to find the initial examples themselves.

Researchers have therefore investigated new tools and protocols for the discovery of relevant bootstrapping examples. These tools often take the form of browsing interfaces whose aim is to help the user exploring the information space in order to locate the sought items. Similarity-based visualization (see *e.g* [15,

* This work is supported by the Swiss National Science Foundation (SNSF) and the EU NoE Petamedia

16]) organizes images with respect to their perceived similarities. Similarity is mapped onto the notion of distance so that a dimension reduction technique may generate a 2D or 3D space representation where images may be organized. A number of similar interfaces have been proposed to apply to the network of users or documents but most browsing operations are based on global hyperlinking (*e.g* Flickr or YouTube pages).

Another popular option is the use of keywords as a mean to apply an initial loose filtering operation over the collection. However, the possibility of responding keyword-based queries depends on the availability of textual annotation over the collection. To be scalable, this option must include a way of making best use of the shallow annotation provided over a subset of documents. Recent data organisation over the WWW, mixing large collections of multimedia documents and user communities offer opportunities to maintain such level of annotation and enable efficient access of information at large scale.

Here, we formalise a model for networked media collections (section 2) that unifies most operations made at the collection level as propagation of information (*e.g* annotation) within a multigraph (section 3). An example of developments exploiting these structures over documents only is presented in section 4.

2 Multidimensional networked data modeling

We start with a collection $\mathcal{C} = \{d_1, \ldots, d_N\}$ of N multimedia items (text, images, audio, video,...). Traditionally, each document d_i may be represented by a set of features describing the properties of the document for that specific characteristic. In a search and retrieval context, it is expected that mainly discriminant characteristics are considered (*i.e* the characteristics that will make it possible to make each document unique w.r.t a given query). With each of these characteristics is associated a similarity measure computed over the document extracted features. Hence, given $\mathcal{C}$, one may form several similarity matrices $S^{[c]} = \left(s_{ij}^{[c]}\right)$, where the value of $s_{ij}^{[c]}$ indicates the level of similarity between documents d_i and d_j w.r.t characteristic c.

In our context, we consider each matrix $S^{[c]}$ as a weighted graph connectivity matrix. Since $S^{[c]}$ is symmetric, it represents the connectivity matrix of a complete non-oriented graph where nodes are documents. Collecting all matrices $S^{[c]}$ $\forall c$, we may therefore represent our collection as a multigraph acting over the node set $\mathcal{C}$.

This simple similarity-based mapping of the collection provides a useful dimensionless representation over which to act in view over efficient collection exploration. In [17], the High-Dimensional Multimodal Embedding (HDME) is presented as a way to preserve cluster information within multimedia collections. In our context, it forms a useful mapping for projection or dimension selection for visualization. It is also a way of enhancing our Collection Guiding principle proposed in [10].

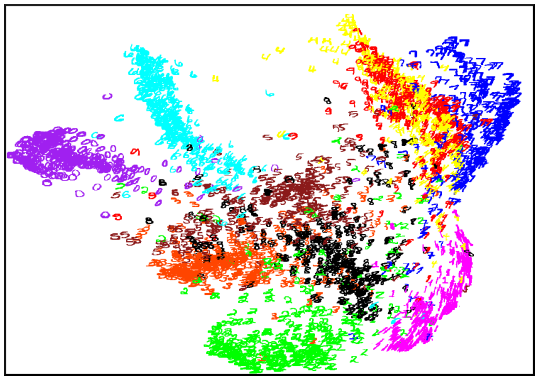


Fig. 1. Global representation of the MNIST handwritten digit image collection after cluster-preserving dimension reduction

Alternatively, documents may be attached a form of metadata taken from a knowledge base $\mathcal{B} = \{b_1, \ldots, b_M\}$ modeled again as a multi-graph over a set of M concepts b_j, acting as nodes bearing relationships between themselves. Distance relationships $D^{[s]} = \left(b_{ij}^{[s]}\right)$ between concepts apply here . The value $b_{ij}^{[s]}$ indicates how much concepts b_i and b_j are close to each other with respect to interpretation s. Examples of such relationships are tags (acting as concepts) whose distance is measured using any word distance measure (*e.g* based on WordNet).

In turn, documents and concepts are also associated with "tagging" relationships $T^{[s]} = \left(t_{ij}^{[s]}\right)$, where $t_{ij}^{[s]}$ evaluates the strength of association between document d_i and concept b_j, under a given perspective (*e.g* interpretation, language) s. Again, the notion of perspective allows for creating multiple relationships between documents and tags (including with respect to users who potentially authored these relationships, see next).

Consider finally a population $\mathcal{P}$ of P users $\mathcal{P} = \{u_1, \ldots, u_P\}$ interacting with the collection $\mathcal{C}$, thus forming a classical social network. Thus, users may be associated by inter-relationships (*e.g* the "social graph" , to use the term coined by FaceBook). Classical relationships such as "is a friend of" or "lives nearby" may be quantified for each pair of users. Matrices $P^{[v]} = \left(p_{kl}^{[v]}\right)$ may thus be formed, where the value of $p_{kl}^{[v]}$ indicates the strength of the proximity aspect v between user u_k and user u_l.

We then consider that any user u_k may have one or more relationships with a document d_i. For example, user u_k may be the *creator* of document d_i or u_k may have *ranked* the document d_i a certain manner. For each of these possible relationships, we are therefore able to form a matrix $R^{[v]} = \left(r_{ik}^{[v]}\right)$, where the value of $r_{ik}^{[v]}$ indicates the strength (or simply the existence) or relation v between

document d_i and user u_k. Users may then be associated with concepts of the knowledge base $\mathcal{B}$ by similar relationships. Essentially, a user may be associated with a concept that describes some particulars of that user (*e.g* being a *student*) or, conversely a concept may be associated by a user because this user has used this concept to annotate documents. Hence, relationships $V^{[s]} = \left(v_{ij}^{[s]}\right)$ are created, where $v_{ik}^{[e]}$ weights a particular link between concept b_i and user u_k.

In summary, we obtain $(\mathcal{C}, S^{[c]}\ \forall c)$, $(\mathcal{B}, D^{[s]}\ \forall s)$ and $(\mathcal{P}, P^{[v]}\ \forall v)$ as multigraphs acting over document, concepts and user node sets and the graph $(G, E) = (\mathcal{C} \cup \mathcal{B} \cup \mathcal{P}, R^{[v]} \cup D^{[s]} \cup V^{[s]})$ as a multi-tripartite graph relating documents, concepts and user node sets. Figure 2 illustrates this representation.

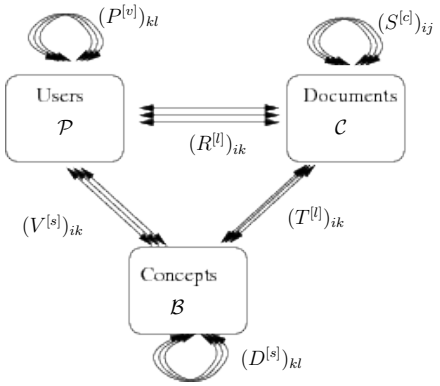


Fig. 2. The proposed graph-based modeling of a social network and associated annotated documents

Now, interestingly, this representation is a base tool for further network analysis and completion. Graph connectivity analysis of $(\mathcal{P}, P^{[v]}\ \forall v)$ for a given aspect v may tell us about coherence between parts of the population. Tools such as minimum vertex- or edge-cuts will indicate particular users or groups that are critical to maintain the connectivity and thus the coherence of the networked information structure.

This structure also allows for its own completion. Similar to what is proposed in [11], user interaction may be captured as one particular bipartite graph $(\mathcal{C} \cup \mathcal{P}, R^{[v]})$ and this information may be mined to enrich either inter-documents similarity $(\mathcal{C}, S^{[c]})$ (see section 3) or inter-user proximity $(\mathcal{P}, P^{[v]})$ to identify a community with specific interests (materialized by the interaction over certain groups of documents). When forming such new relationships, constraints for

forming proper distance matrices (1:a) or similarity matrices (1:b) apply:

$$(a)\begin{cases} s_{ij}^{[c]} \geq 0 \\ s_{ij}^{[c]} = s_{ji}^{[c]} \\ s_{ii}^{[c]} = 0 \end{cases} (b)\begin{cases} 0 \leq s_{ij}^{[c]} \leq 1 & \forall i,j \\ s_{ij}^{[c]} = s_{ji}^{[c]} & \forall i,j \\ s_{ii}^{[c]} = 1 & \forall i \end{cases} \quad (1)$$

Similarly, recommender systems [7] will mine inter-user relationships and inter-document similarity to recommend user-document connections.

3 User relevance modeling

We are first interested in exploiting the graph G to mine a posteriori usage of interactive information systems, with the aim of using user interaction as a source of semantic knowledge that will enrich the descriptions of documents. In other words, with reference to our model in Figure 2, we will mine relationships $R^{[v]}$ (user-documents), in order to enhance the understanding of the structure of $T^{[s]}$ (documents-concepts) and $S^{[c]}$ (documents-documents).

Given a query-by-example retrieval system that affords relevance feedback, we can assume that, at any given stage, users will have invoked a set of L queries $\mathcal{Q} = \{q_1, ..., q_L\}$ over the set of documents $\mathcal{C}$. An $N \times L$ document-query relevance matrix $\mathcal{R}$ can then be defined, where each element

$$\mathcal{R}(i,j) = \begin{cases} +1 & \text{if the user marked document } d_i \text{ as positive in query } q_j, \\ -1 & \text{if the user marked document } d_i \text{ as negative in query } q_j, \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

The relevance matrix $\mathcal{R}$ thus formed is an instantiation of matrix $R^{[v]}$ defined in section 2. The proposed User Relevance Model formalizes a missing aspect of previous studies in long-term learning. Assumptions made in previous studies to generate and model artificial data ignore the concept-basis relationship between a document and a query, noise introduced due to user error, and oversimplify the user decision in judgment-making [14, 6, 2]. This new formalization permits the understanding of how long-term RF data is generated by users based on perceived concepts in the documents and queries.

The idea is then to examine the RF process to discover the underlying concepts present in the documents and queries. Essentially, we want to discover to what extent each concept $b_i \in \mathcal{B}$ exists in $d_i \in \mathcal{C}$ and $q_j \in \mathcal{Q}$.

Generally, decompositions made by latent-variable models are not unique and therefore the interpretation of the latent variables can be problematic [1]. However, the latent space present in the component matrices can be interpreted in light of the values of the rows and columns in the co-occurrence matrix.

For example, consider two images d_1 and d_2 depicting horses. Through a decomposition both documents are seen to have a high component of concept b_1. In the absence of further information, we could say that concept b_1 may represent something to do with horses.

Non-negative matrix factorization (NMF) offers a straightforward approach to the problem of discovering latent concepts from observed data. NMF, given a non-negative matrix $\mathcal{R}$, finds non-negative, non-unique factors giving:

$$\mathcal{R} \approx WH, \tag{3}$$

where $W \in R^{M \times K}$ and $H \in R^{K \times N}$ and such that $W \cdot H$ minimizes the Frobenius norm $||\mathcal{R} - WH||^2$ [9]. In our case, the resulting component matrix W yields a projection of the documents into the space defined by the latent basis vectors.

Another popular method is latent semantic analysis (LSA) which uses the singular value decomposition (SVD) to decompose the co-occurrence matrix into three components:

$$\mathcal{R} = U \Sigma V^T, \tag{4}$$

where U are the left singular vectors, Σ are the singular values, and V are the right singular vectors [4]. The decomposition is such that U and V are orthonormal, and Σ is a diagonal scaling matrix with values in decreasing order. By retaining only the first K singular values in Σ, we have an rank-K approximation of the original co-occurrence matrix:

$$\mathcal{R}_K \approx U_K \Sigma_K V_K^T, \tag{5}$$

where $U_K \in R^{M \times K}$, $\Sigma_K \in R^{K \times K}$, and $V_K \in R^{N \times K}$.

In choosing the number of latent variables appropriately in NMF and the SVD, we can conveniently extract the underlying concepts in the User Relevance Model. We can identify the concept weight matrices $T^{[s]} = \left(t_{ij}^{[s]}\right)$ as having the same dimensionality as W, U_K and H, V_K^T, respectively. In practice, we will never know the exact nature of the underlying concepts, and so M must be chosen such that it is large enough to capture diversity in the data yet small enough that some interpretable clustering is observed.

Because the singular value decomposition does not impose non-negativity assumptions on the co-occurrence matrix or the resulting component matrices, interpretation of the latent variables in U_K and V_K^T is not straightforward [8]. NMF is preferable in this sense, because the latent variables lend themselves to a modeling of non-negative concept weights, which we note leads to a probabilistic formulation [5].

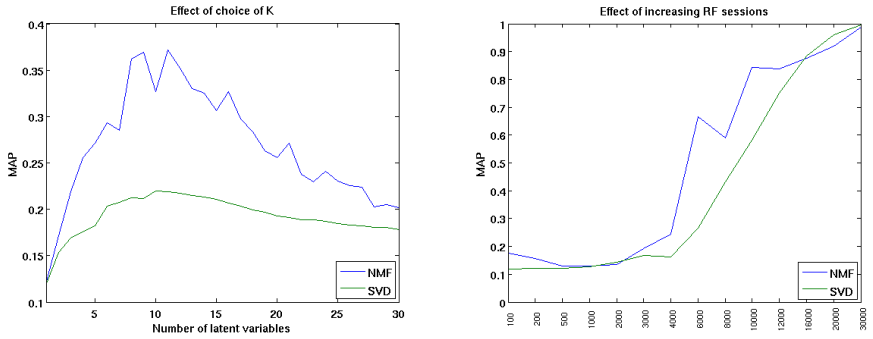
3.1 Experiments

Illustrative experiments are conducted on a small subset of the Corel image collection (see also [13]). The subset comprises 1,000 images uniformly spanning 10 categories (100 images per category). Although small, this dataset allows us to quickly and easily visualize performance. Document categories are contiguous in the matrix $\mathcal{R}$ and therefore similarly in all figures to make interpretation of the results easier.

All sessions of relevance feedback are generated according to the User Relevance Model described above. In other words, given the ground truth image

categories, we generate a full relevance matrix and subsequently account for real-world sparsity and noise, yielding $\mathcal{R}$. Performance is measured using mean average precision (MAP). MAP emphasizes retrieving relevant documents first and provides a quantifiable measure of the clustering of the documents into latent concept classes.

The subplots of Figure 3 show the effects of varying various parameters in the User Relevance Model.



(a) Choice of number of latent variables (K)

(b) Number of RF sessions

Fig. 3. Parameters of the User Relevance Model are varied to show the MAP under different conditions.

We know from previous work that the MAP should be at its maximum when the value of K is equal to the actual number of concepts underlying the data [12]. In Figure 3 (A), MAP is highest when $K \approx 10$. Figure 3 (B) demonstrates that the MAP increases as we collect more relevance feedback judgments (sessions). It is evident that the MAP approaches 1 as the number of relevance feedback sessions ML increases. A significant improvement in mean average precision is observed with as little as 6,000 RF sessions.

Figure 4 demonstrates the success of NMF's modeling of the concept weights from the User Relevance Model. The figure shows the highest ranked images for the particular concept. Figure 5 shows the reconstructed concept matrix W containing weights for each document. Columns corresponds to the bar plots in Figure 4. Due to the inherent non-uniqueness of latent-variable models, the columns of W and the original document-concept matrix will not correspond. What is important to note is that the documents are grouped into similar clusters.

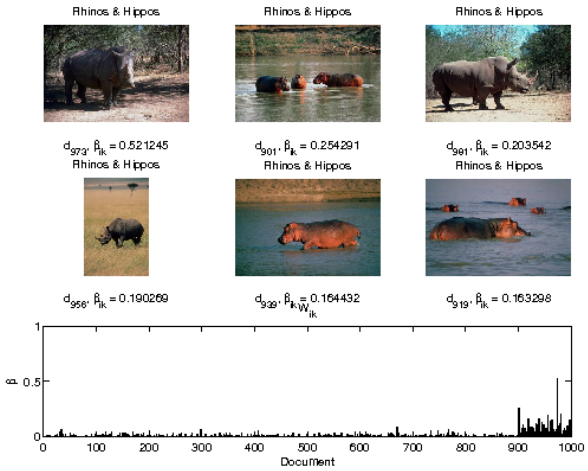


Fig. 4. Example underlying concept (bottom) with corresponding images depicting images of rhinoceri and hippopotami. Beneath each document is the corresponding concept weight.

4 Multidimensional data exploration

We are now interested in defining exploration strategies over social networks. Based on the above modeling, we map this challenge onto that of defining optimal discrete structures in the multigraphs representing the social network. The objective is to complement the search paradigm with a navigation facility. We therefore assume that a search tool is used to position a user (called *client* to differentiate from users in the network) at a certain point within our multigraph by selecting a particular user or a particular document. From that point on, the navigation system should enable the client to move within a neighborhood, as defined by the connectivity structure, to explore the vicinity of this position. In other words, we wish to offer the client a view of where to navigate next and this view should be optimized from the information available. Further, this recommendation should be embedded within a global context so as to avoid cycles where the client stays stuck within a loop in the navigation path.

Our graph model is a suitable setup for this optimization. Formally, starting with a matrix $M = (m_{ij})$, where m_{ij} indicated the *cost* of navigating from item x_i to item x_j, we wish to find a column ordering o^* of that matrix that will minimize a certain criteria over the traversal of the items in this order. As a basis, we seek the optimal path that will minimize the global sum of the costs associated to the traversed edges. That is, we seek o^* as the ordering that will minimize the sum of the values above the diagonal so that

$$o^* = \arg\min_{o} \sum_{i \in o} m_{ii+1} \tag{6}$$

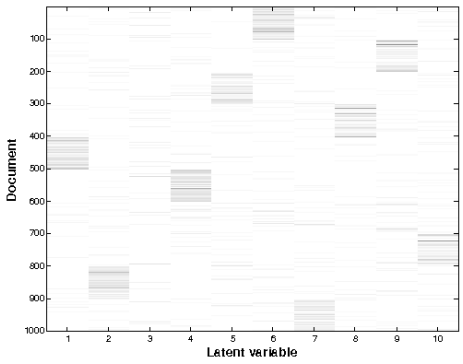


Fig. 5. Recovered document-concept relationships in W. Documents are clustered into the latent concepts. The concept in column 3 shows overlap between images. This is attributed to noise introduced in the User Relevance Model.

The above is equivalent to solving the Symmetric Travelling Salesman Problem (S-TSP) over the complete graph with arc cost m_{ij}. The tour thus forms an optimal discrete structure to explore the complete set of nodes while minimizing the sum of the lengths of the edges traversed during the tour.

In our model, matrices $1 - S^{[c]} = \left(1 - s_{ij}^{[c]}\right)$ and $1 - P^{[v]} = \left(1 - p_{kl}^{[v]}\right)$ follow constraints (1:a) and are therefore suitable inputs for the S-TSP procedure. Figure 6 illustrates the effect of column-reordering on a set distance matrix. The values over the diagonal m_{ii+1} are taken as step costs during the navigation and their overall sum is minimized. We have applied the above principle over

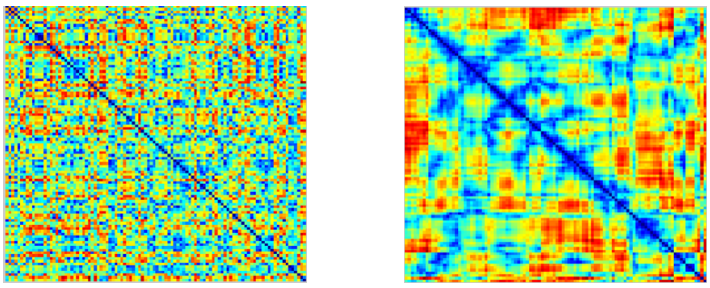


Fig. 6. An optimal reordering applied over a distance matrix (color similarity between 300 images)

a collection of Cultural Heritage digital items composed of images (paintings, historical photographs, pictures,...) annotated with description and metadata. We have defined several browsing dimensions using visual, temporal and textual

characteristics. The result is a Web interface presenting a horizontal browsing dimension as its bottom line. From each of these items, a complementary vertical path is displayed. Image size is used to represent the distance from the main focal item displayed at the center of the bottom line. In essence, our browser uses a strategy closes to that implemented in [3].

Fig. 7. The proposed browsing interface

Figure 7 shows the interface with the focal point in the bottom blue box. Green vertical and horizontal arrows materialize the paths that may be followed. In the upper right corner (dashed red box), a table displays the summary sample of the collection of items. Clicking on any of these images (but the central one) brings it to the center and updates its context (*i.e* computed neighborhood). Clicking on the central image goes back to the search interface as a complement to the navigation mode and displays the full details (*e.g* metadata) of this particular document. The choice of tours followed along the horizontal and vertical axis is regulated by setting options at each step of the browsing, thus allowing "rotations" around the focal point to display any combination of dimensions.

Concerning the modeling of potential attached social network ($\mathcal{P}$), we are planning to include an interface enabling the browsing of population formed by the creators of the documents, in parallel with this document browsing interface. We have applied the very same principle with different features to browse meeting slide collections with respect to visual slide similarity (to identify reuse of graphical material), textual similarity (to relate presentations by topic and timeline (to simply browse thru the presentation). Again, a social of presentation authors may complement this document browsing tool.

5 Conclusion and discussion

In this paper, we propose a modeling of data immersed into a social network based on multigraphs. We show how these multigraphs may be the base for

defining optimal strategies for information discovery in complex and large documents collections, based on exploiting user interaction. Latent (topic) models are promising tools in this context. The choice of modeling may impose a particular model, based on the ease on interpretation of its parameters.

We also demonstrate that, via the definition of optimal structures, efficient exploitation of the network structure and contained data may be achieved. In particular, we advocate the use of the S-TSP for organizing a unique navigation path to be followed. Many graph-based structures are defined by NP-Complete problems. The problem of scalability thus forces proper approximations to be found.

Our work has been evaluated in all the steps of its engineering (*e.g* similarity computation). However, we still must complete our evaluation by the final usability of the produced interfaces. Initial demonstration sessions with credible tasks show encouraging interests from various classes of users (clients). Only quantitative tests over well-defined tasks and measures will tell us how much these interfaces are actually able to complement classical query-based search operations.

References

1. D. J. Bartholomew and M. Knott. *Latent variable models and factor analysis.* Oxford University Press, Inc., New York, NY, USA, 1999.
2. M. Cord and P. H. Gosselin. Image retrieval using long-term semantic learning. In *IEEE International Conference on Image Processing*, 2006.
3. Scott Craver, Boon-Lock Yeo and Minerva Yeung. Multi-linearisation data structure for image browsing. In *SPIE Conf. on Storage and Retrieval for Image and Video DBs VII*, 1999.
4. S. Deerwester, S. Dumais, T. Landauer, G. Furnas, and R. Harshman. Indexing by latent semantic analysis. *Journal of the American Society of Information Science*, 4:391–407, 1990.
5. Eric Gaussier and Cyril Goutte. Relation between plsa and nmf and implications. In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 601–602, New York, NY, USA, 2005. ACM.
6. Xiaofei He, O. King, Wei-Ying Ma, Mingjing Li, and Hong-Jiang Zhang. Learning a semantic space from user's relevance feedback for image retrieval. *Circuits and Systems for Video Technology, IEEE Transactions on*, 13(1):39–48, 2003.
7. Thomas Hofmann. Latent semantic models for collaborative filtering. *ACM Transactions on Information Systems (TOIS)*, 22(1):89 – 115, 2004.
8. A. Kabán and M. A. Girolami. Fast extraction of semantic features from a latent semantic indexed text corpus. *Neural Process. Lett.*, 15(1):31–43, 2002.
9. D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, October 1999.
10. Stéphane Marchand-Maillet and Éric Bruno. Collection guiding: A new framework for handling large multimediacollections. In *Audio-visual Content And Information Visualization In Digital Librairies*, Cortona, Italy, 2005.

11. D. Morrison, E. Bruno and S. Marchand-Maillet. Capturing the semantics of user interaction: A review and case study. In *Emergent Web Intelligence*. Springer, 2010.
12. Donn Morrison, Stéphane Marchand-Maillet and Eric Bruno. Semantic clustering of images using patterns of relevance feedback. In *Proceedings of the 6th International Workshop on Content-based Multimedia Indexing*, London, UK, June 18-20 2008.
13. Donn Morrison, Stéphane Marchand-Maillet and Eric Bruno. Modelling long-term relevance feedback. In *Proceedings of the ECIR Workshop on Information Retrieval over Social Networks*, Toulouse, FR, April 6th 2009.
14. Henning Müller, Wolfgang Müller, David McG. Squire, Stéphane Marchand-Maillet, and Thierry Pun. Long-term learning from user behavior in content-based image retrieval. Technical report, Université de Genève, 2000.
15. G.P. Nguyen and M. Worring. Optimization of interactive visual similarity based search. *ACM TOMCCAP*, 4(1), 2008.
16. Y. Rubner. *Perceptual Metrics for Image Database Navigation*. PhD thesis, Stanford University, 1999.
17. E. Szekely, E. Bruno and S. Marchand-Maillet. High dimensional multimodal embedding for cluster preservation. Technical Report VGTR:0801, Viper - University of Geneva, 2008.