A Graphical Model for Context-Aware Visual Content Recommendation

Article *in* IEEE Transactions on Multimedia · February 2008

DOI: 10.1109/TMM.2007.911226 · Source: IEEE Xplore

A Graphical Model for Context-Aware Visual Content Recommendation

Sabri Boutemedjet and Djemel Ziou

***Abstract*—Existing recommender systems provide an elegant solution to the information overload in current digital libraries such as the Internet archive. Nowadays, the sensors that capture the user's contextual information such as the location and time are become available and have raised a need to personalize recommendations for each user according to his/her changing needs in different contexts. In addition, visual documents have richer textual and visual information that was not exploited by existing recommender systems. In this paper, we propose a new framework for context-aware recommendation of visual documents by modeling the user needs, the context and also the visual document collection together in a unified model. We address also the user's need for diversified recommendations. Our pilot study showed the merits of our approach in content based image retrieval.**

***Index Terms*—Collaborative filtering, content-based image retrieval, content-based image suggestion, context-aware retrieval, diversity ranking, image summarizing, mixture models, recommender systems.**

I. Introduction

THERE is a huge amount of digital data produced every day in the World Wide Web. This digital data take different forms such as text, sound, images and videos. Information retrieval (IR) provides tools and techniques that help users to access, browse, summarize that information efficiently. In the case of visual information, these techniques are addressed within content based image retrieval (CBIR) community. In retrieval, a user expresses the information need by formulating a search query generally in the form of image examples or textual descriptions. Then, the CBIR system retrieves from the collection those visual documents that are close to the user's query. The kind of information needs addressed in CBIR is short term. The goal is to respond to the user's search query. There is another kind of interests i.e., long term or permanent such as desires, tastes and preferences of each user. For example, in a marketing domain, visual documents have been recognized as efficient means in advertisements since they can convey meanings that cannot be expressed using words [1]. Indeed, managing long term user interests to visual information is crucial. To that end, we propose the Content-Based Image Suggestion (CBIS) to address the modeling of user preferences with respect to visual information.

Research in consumer psychology has shown that the product adoption is synonymous to the willing to acquire novel information [2] and is influenced by preferences and the external environment [3], [4] such as the time and place. Further, the consumers seek for newness in their choices in order to maintain the stimuli at the optimum levels [2] and for diversity to build hedges against uncertainty in their future tastes [5]. Both behaviors are referred to as consumer's novelty seeking and variety seeking behaviors, respectively. In a consumer's decision making, these behaviors compete with the user's conformity to social norms i.e., similarity with others purchase history and was explained by the user's need to get rewards from others.

Recommender systems [6] employ information filtering (IF) technologies in order to satisfy user needs caused mainly by user's conformity. On the other hand, context-aware information retrieval has emerged recently [7] and try to model user's short term needs influenced by the contextual information such as time, place or even handheld device characteristics (low computing power, screen resolution and telecommunication cost) in a mobile computing environment.

Let us illustrate the CBIS problem by the example in Fig. 1 which shows the history of two users John and Mary with similar preferences during two different contexts $\{\text{weekday}, \text{weekend}\}$. We would like to suggest novel images of mountains from the collection to John during weekends which is not possible to do using existing IF techniques. Further, the recommender Viscors [8] which uses Pearson Correlation Coefficients approach [9] cannot compute the correlation between Mary and John since there is no visual document rated by both users in their history.

The CBIS system we propose in this paper inspires from consumer psychology researches by considering the following factors in order to make recommendations of visual documents: (1) the visual content and metadata and (2) user preferences. The visual content can be described by keywords and visual features extracted form the documents. These features may be local, global, low level or of semantic nature [10]. The keywords can be automatically or semi-automatically extracted by annotation or recognition process. The user preferences are predicted by considering the user's conformity, the external environment (i.e., context) and the user's novelty-seeking and variety-seeking behaviors. The user's conformity is predicted by considering the history of a community of similar or "like-minded" users. On the other hand, the user variety-seeking and novelty-seeking behaviors are met by using the visual and semantic similarity among visual documents of the purchase history and recommendation list, respectively.

Manuscript received August 10, 2006; revised August 21, 2007. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and Bell Canada through its Bell University Laboratories R&D Program. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jie Yang.

The authors are with the Département d'Informatique, Universitéde Sherbrooke, Sherbrooke, QC J1K 2R1 Canada (e-mail: sabri.boutemedjet@usherbrooke.ca; Djemel.ziou@usherbrooke.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2007.911226

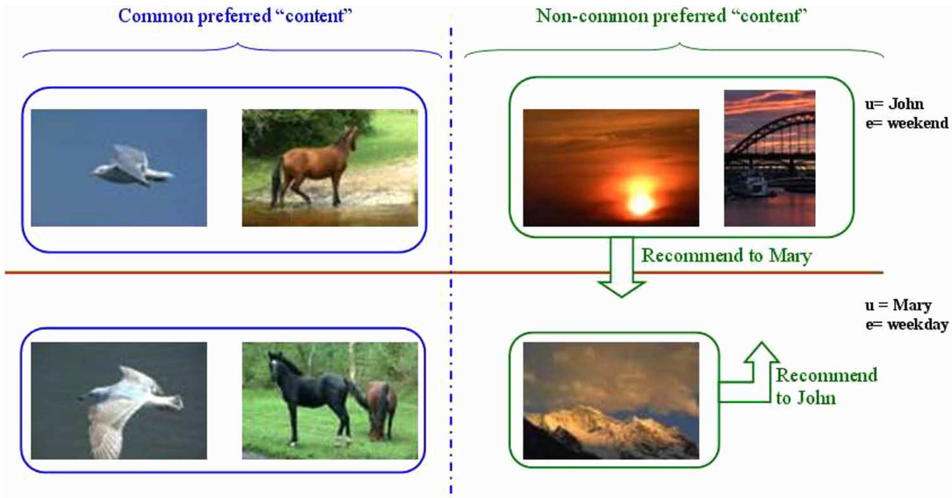


Fig. 1. Principle of content-based image suggestion.

This paper is organized as follows. Next, in Section II, we present the state-of-the-art of recommender systems. Then, in Section III, we propose a graphical model for CBIS and we estimate its parameters in Section IV using a variational inference. After that, we detail a unified approach for handling the diversity of recommendation lists in Section V. Then, we analyze the computational complexity of the proposed method in Section VI. The experimental results are presented in Section VII. Finally, we conclude this paper by a summary of the work.

II. Related Research Work

In information filtering (IF), there exist in literature three families namely content-based filtering (CBF), collaborative filtering (CF) and hybrid methods. CBF employs information retrieval techniques in representing user profiles using content descriptors [11]. For example, [12] employs a Multinomial text model learned for each user from information about books he or she has seen in the past. CBF suffers from the overspecialization problem since they cannot recommend to the active user "novel" and "unexpected" items different from his/her profile. On the other hand, CF techniques identify the neighbors or "like-minded users" of the active user. We distinguish two kinds of algorithms in CF: memory-based approaches [13] and model-based approaches. In memory-based approaches such as the Pearson Correlation Coefficients (PCC) [9], the prediction of the rating for the active user is made on the basis of the ratings of other users with similar interests. Model-based CF techniques [14] learn first a statistical model of user/item classes (e.g., clustering) and then predict the ratings based on the learned model. Examples of CF models include the Bayesian Clustering (BC) [15], the Aspect model [16], the Flexible mixture model (FMM) [17] and many others. Bayesian clustering assumes the availability of K user classes each of them will rate items similarly. The Aspect model [16] associates different states of a latent variable to different pairs (i.e., co-occurrence) of users and items which become conditionally independent given the state of the latent variable. The

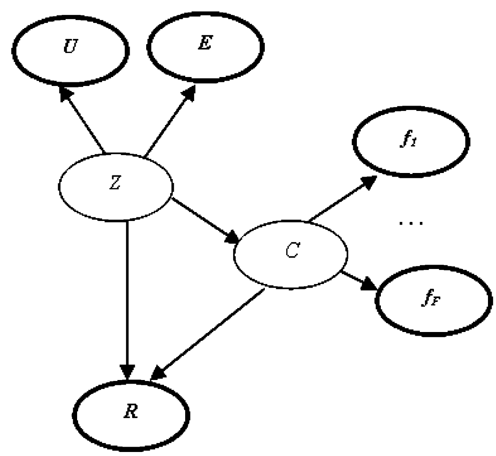


Fig. 2. Graphical representation of the VCAF model.

Flexible Mixture Model (FMM) [17] goes further in flexibility by introducing two latent variables to cluster separately both users and items. From the fact that items are considered as a categorical variable (i.e., unique index for each item) within the CF framework, then the recommender cannot suggest nonrated items. In addition, CF models overfit under high data sparsity levels e.g., when only very few ratings are available [18], [19]. To alleviate these problems hybrid methods were proposed and use the item content to improve the rating prediction. For example, the authors in [19], generate new pseudo-ratings based on word features associated with movies to fill out the user-item matrix (data set). Another hybrid model is proposed in [20] which refines the item class memberships of the FMM by training a separate discriminative model for item classes given word features. The new item class parameters are then used in the generative model to better fit user and item classes. We notice, however, that hybrid filtering approaches have been used mainly for textual data.

Our work contributes by the following. First, the added value of the visual and textual information of each visual document is modeled to improve the user's satisfaction. Second, we provide

Fig. 3. Left: diversity-based recommendation and right: rating-based recommendation.

more personalization by handling changing user preferences according to different contextual situations. Finally, visual documents are selected according to both their relevance to the user (i.e., conformity) and the diversity they provide with respect to the already suggested items.

III. Proposed Model

We consider the problem of CBIS as the maximization of a certain utility that ranks visual documents for a user in a certain context. The domains we consider consist of a set of users $\mathcal{U} = \{u_1, u_2, \ldots, u_{N_U}\}$, a set of visual documents $\mathcal{X}$, and a set of possible contexts $\mathcal{E} = \{e_1, e_2, \ldots, e_{N_E}\}$. The relevance is expressed explicitly on an ordered voting i.e., rating scale defined as $\mathcal{R} = \{r_1, r_2, \ldots, r_{N_R}\}$. For example, the binary scale $\{-1, +1\}$ can be used to express "rejection" or "acceptance" preferences while the five star scale (i.e., $N_R = 5$) such as the one used by Amazon (http://www.amazon.com) allows the users to give more detailed degrees of appreciation. In order to meet user needs in diversity, the recommendation list of relevant visual documents should be diversified [2], [21]. The example illustrated by Fig. 3 shows clearly that the left recommendation list would be preferred to the right one since it contains diverse content.

Let $\mathcal{X}^{ue} = \{x_1^{ue}, x_2^{ue}, \ldots, x_L^{ue}\}$ be a list of L ranked visual documents to recommend to a given user u in a context e where $x_i^{ue} \in \mathcal{X}, i = 1, \ldots, L$. We denote by x_t^{ue} the visual document at the rank t in $\mathcal{X}^{ue}$. The users can be mobile and receive recommendations on either PDAs (Portable Digital Assistants) for which the size of $\mathcal{X}^{ue}$ is generally small i.e., $L < 10$ or desktop computers i.e., $L < 25$. We notice that the diversity of $\mathcal{X}^{ue}$ imposes another condition that involves measuring dependencies (information redundancies) within subsets of visual documents during the suggestion process. It has been argued in IR [22], [23] that the selection of a set of "relevant" and "diversified" documents is an NP-hard problem for which greedy optimizations provide locally optimal solutions.

Following a similar methodology of IR, we propose to fill the recommendation list $\mathcal{X}^{ue}$ by selecting successively x_t^{ue} that are relevant (i.e., vote) and novel. The degree of novelty of x_t^{ue} is generally measured in terms of the information redundancy it provides with respect to a certain set of visual documents $\mathcal{X}_t^{ue}$. Thus, we propose a utility $s(x, u, e|\mathcal{X}_t^{ue})$ which computes a score for a visual document x according to its relevance for the user u in the context e given $\mathcal{X}_t^{ue}$. Then, the Visual document Context Aware Filtering (VCAF) model underlying CBIS is given in the following:

$$x_t^{ue} = \arg\max_{x \in \mathcal{X}} s(x, u, e|\mathcal{X}_t^{ue}). \tag{1}$$

A diversified $\mathcal{X}^{ue}$ can be obtained by considering $\mathcal{X}_t^{ue} = \{x_1^{ue}, \ldots, x_{t-1}^{ue}\}$. One would like to consider also the novelty w.r.t to a set $\mathcal{D}_T^{ue}$ of delivered visual documents during a window of time T such as a week or a month [21], then, $\mathcal{X}_t^{ue} = \mathcal{D}_T^{ue} \cup \{x_1^{ue}, \ldots, x_{t-1}^{ue}\}$. Finally, if the diversity of $\mathcal{X}^{ue}$ is not of interest, then, we have an independence between $\mathcal{X}_t^{ue}$ and x_t^{ue}. Therefore, the appropriate model for the suggestion would be $s(x, u, e|\mathcal{X}_t^{ue}) = s(x, u, e)$. Next, we will first focus on modeling the data underlying the suggestion, then, we will show in Section V how this data model is used to make recommendations.

A. Data Modeling

Many utility functions for model-based approaches were proposed in literature. They depend mainly on the goal of the information filter and the representation of items. For the goal "*recommend N relevant items*," the common utility employed by CF is the predicted rating through $p(r|d, u)$ [16], where d is the index of the item. The authors in [17] model the joint event $p(u,d,r)$ and then predict the rating using the mean rule as $\sum_r rp(r|d,u)$ where the quantities $p(r|d,u)$ are derived by conditioning $p(u,d,r)$ on (d,u). Also, in [24], the positive and negative ratings are modeled through $p(u,d,r^+)$ and $p(u,d,r^-)$ and items are ranked by their Bayes' factor $p(r^+|u,d)/p(r^-|u,d)$. Similarly, CBF techniques [12] define a model for classes of liked c_0 and disliked c_1 documents based on text features. Then, documents are ranked by their odds ratio $p(c_1|x)/p(c_0|x)$. From the document representation point of view, the content features used for x will depend on the target application domain. In remote sensing for example, one could combine the context (date, geographic position, device or sensor) with visual and/or semantic features in describing visual documents [25]. By this way, we capture user preferences related to different kinds of content representation such as visual, textual, and contextual. In addition, from context-aware information retrieval point of view [7], these user preferences depend also on each user's context. Therefore, we propose a unifying model which integrates seamlessly information about users, their contexts and content features. By considering a visual document as a set of F feature vectors f_i i.e., $x=(f_1,\ldots,f_F)$, its relevance for a given user u and context e, can be evaluated by *"extending"* one of the state-of-the-art utilities e.g., $p(r^+|x,u,e)/p(r^-|x,u,e)$. We employ a generative approach similarly to CF techniques [14] and we model the joint event $p(u,e,x,r)$. By this way, one could deduce other utilities such as the one required in a alerting system [11] (e.g., surveillance) which *"alerts"* an operator by a certain captured event x. Such scenarios are achieved through $p(u,e|x,r^+)$ that can be derived from $p(u,e,x,r)$. An interesting discussion about generative graphical models for CF can be found in [14].

In order to generalize for unseen $< u,e,x,r >$, we propose to introduce two latent variables c and z. The variable c denotes visual content classes and organizes the image collection by grouping similar contents into homogeneous clusters. On the other hand, the variable z denotes user classes and explains the user's conformity by modeling the similarity of user preferences. Indeed, the model is flexible since a user may belong to many classes with different probabilities. By considering K states for z and M classes of visual content, we can develop the generative model $p(u,e,x,r)$ by standard marginalization over z and c as

$$p(u,e,x,r)=\sum_{c=1}^{M}\sum_{z=1}^{K}p(u,e,x,r,z,c). \tag{2}$$

The "complete" factorization (without assumptions) of $p(u,e,x,r,z,c)$ using the chain rule leads to densities with a huge number of parameters which are difficult to interpret in terms of the data. For this reason, we need to assume some conditional independencies among variables that represent the effect of causality [26] motivated from either consumer psychology or information management. We assume that we are interested by users and contexts jointly since the recommendation in general is initiated for a given user who belongs to a certain context. This constitutes one main difference with existing IF approaches which do not consider the influence of the context [4] on the user preferences. We assume that given the knowledge of the user class z makes (u,e) conditionally independent from (x,r,c). This corresponds to the fact that z *"summarizes"* any information about (u,e). From graphical models theory [26], if two random variables A and B are conditionally independent given the random variable C, then, $P(A|B,C)=P(A|C)$. Indeed, by assuming that a visual document is influenced only by its class c, then we have: $p(x,r,c|z)=p(c|z)p(r|z,c)p(x|c)$. This assumption is motivated by the fact that c "summarizes" any information about x. From previous developments, we give our generative model in the following:

$$p(u,e,x,r)=\sum_{c=1}^{M}\sum_{z=1}^{K}p(z)p(u,e|z)p(c|z)p(r|z,c)p(x|c). \tag{3}$$

We recall that the model given by (3) can be seen as a pure CF by using unique resource locators (url) $d=x$ as descriptors. For example, considering $\mathcal{E}$ as a singleton, then, (3) can be seen as the FMM model [17]. Moreover, by considering $x=(f_1,\ldots,f_F)$ (hybrid IF), we employ the "bag-of-features" approach which assumes a realistic conditional independence among visual document descriptors $f_l, l=1,\ldots,f$ given the class c. By this way, the modeling task would focus on choosing appropriate models (Multinomial, Gaussian, Dirichlet, etc.) for each descriptor f_l (possibly a vector). Therefore, the VCAF model can be developed as

$$p(x,r,u,e)=\sum_{c=1}^{M}\sum_{z=1}^{K}p(z)p(u,e|z)p(c|z)p(r|z,c) \times\prod_{l=1}^{F}p(f_l|c). \tag{4}$$

The graphical representation of this model is illustrated by Fig. 2 in which nodes denote random variables and edges denote dependencies among variables.

Let us consider the following example by defining the different quantities in the right side of (4). Without loss of generality, we consider in this work two content descriptors for each visual document as $x=(d,\vec{v})$ (i.e., $F=2$) where $\vec{v}$ denotes the feature vector (texture, local features, shape or keywords) extracted from the visual document x and d its index which is categorical. We will have the quantities $p(d|c)$ and $p(\vec{v}|c)$ that appear in the right side of (4). The class conditional distributions $p(\vec{v}|c)$ of visual features are considered Dirichlet distributions (DD) [27] which have proven to be efficient for modeling non-Gaussian data. In fact, the DD is defined in a compact support and offers a higher flexibility due to the possibility to have different shapes. A visual document feature vector $\vec{v}=(v_1,\ldots,v_n)$ follows a DD of the class c with parameters $\vec{\alpha}^c=(\alpha_1^c,\ldots,\alpha_n^c), \alpha_l^c>0$ when $p(\vec{v}|c)$ has the form of the density given in (5). This density is defined on the simplex $\{(v_1,\ldots,v_n),\sum_{l=1}^{n}v_l=A\}$ [28]

$$p(\vec{v}|c)=\frac{\Gamma\left(\sum_{l=1}^{n}\alpha_l^c\right)}{A^{[(\sum_{l=1}^{n}\alpha_l^c)-1]}\prod_{l=1}^{n}\Gamma(\alpha_l^c)}\prod_{l=1}^{n}(v_l)^{\alpha_l^c-1}. \tag{5}$$

Let us now focus on the definition of the remaining quantities in (4). First, $p(u,e|z)$ is the likelihood of a user and context to belong to the class z. They can be initialized from the knowledge about the membership of users and contexts to user classes. This knowledge can be set initially according to some information about user preferences, or by employing an unsupervised clustering technique such as K-Means, or even randomly. Second, $p(z)$ is the class prior for the user class z and can be initialized as the proportion of users and contexts with class membership z. Third, the quantity $p(d|c)$ is the likelihood of a visual document to belong to the content class c. It can be initialized in a similar way to $p(u,e|z)$ or deduced from the likelihood of visual features $p(\vec{v}|c)$ as the follows. We use the fact $p(c|d) \simeq p(c|\vec{v})$ which leads to $p(d|c) = p(d)p(\vec{v}|c)/p(\vec{v})$ where $p(d)$ is the proportion of ratings given to the document x in the data set. Fourth, the quantity $p(c|z)$ denotes the probability that a user class "selects" an image class. It can be initialized from frequencies of labeled observations (with z and c) for a fixed z. Finally, $p(r|z,c)$ informs about probability of a rating r to be generated for a given user class and content class. It can be initialized as the proportion of labeled observations having the rating r for a given z and c.

We notice that all the random variables, r, d, z, and c involved in (4) are discrete. We employ the following notation to simplify the presentation. The probability $p(a|\pi)$ of discrete random variable A conditioned on its parent (predecessor) Π is denoted by $\theta^A_{\pi a}$. By considering $\{1, 2, \ldots, N_A\}$ as the definition domain of the random variable A, then we put $\vec{\theta}^A_\pi = (\theta^A_{\pi 1}, \ldots, \theta^A_{\pi N_A})$ under the constraint: $\sum_{a=1}^{N_A} \theta^A_{\pi a} = 1$. Therefore, from Fig. 2, we have to estimate Θ, the set of all parameters given by

$$\Theta = \left(\vec{\theta}^U_z, \theta^E_z, \vec{\theta}^Z, \vec{\theta}^C_z, \vec{\theta}^R_{zc}, \vec{\theta}^D_c, \vec{\alpha}^c\right). \quad (6)$$

For an unrated ("novel") visual document $x_{\text{new}} = (d_{\text{new}}, \vec{v}_{\text{new}})$ for which $p(d_{\text{new}}|c)$ is not known, we identify its class membership from the information carried out by its feature vector $\vec{v}_{\text{new}}$. Therefore, we consider $p(d_{\text{new}}|c)$ as a constant for all c which leads to the following approximation:

$$p(x_{\text{new}}, u, e, r) \propto \sum_{c=1}^{M} \sum_{z=1}^{K} p(z)p(u,e|z)p(c|z)p(r|z,c) \times p(\vec{v}_{\text{new}}|c). \quad (7)$$

We notice that (6) does not involve updating the parameters of the VCAF model for "unrated" visual documents.

IV. Variational Inference of the Generative Model

We develop in this section the method used to estimate the parameters Θ given by (6). Since the VCAF-model contains hidden variables, then the Expectation-Maximization (EM) [29] algorithm is commonly used for maximum likelihood estimation. We assume a data set of N independent and identically distributed (iid) observations $\mathcal{D} = \{< u^{(i)}, e^{(i)}, x^{(i)}, r^{(i)} > | i = 1, \ldots, N, u^{(i)} \in \mathcal{U}, e^{(i)} \in \mathcal{E}, x^{(i)} \in \mathcal{X}, r^{(i)} \in \mathcal{R}\}$. From [30], the log-likelihood $\log p(\mathcal{D}|\Theta)$ admits the lower bound $\mathcal{F}(Q,\Theta)$ w.r.t. Θ given by (8) obtained using the Jensen's inequality from the fact that the logarithm is concave

$$\mathcal{F}(Q,\Theta) = \sum_{i=1}^{N}\sum_{z=1}^{K}\sum_{c=1}^{M} Q^{(i)}_{zc} \log p(x^{(i)}, r^{(i)}, u^{(i)}, e^{(i)}, z, c) - \sum_{i=1}^{N}\sum_{z=1}^{K}\sum_{c=1}^{M} Q^{(i)}_{zc} \log Q^{(i)}_{zc} \quad (8)$$

where $Q^{(i)}_{zc} = Q(z,c|u^{(i)}, e^{(i)}, x^{(i)}, r^{(i)}; \Theta)$, Q is the *variational probability function* and $p(u^{(i)}, e^{(i)}, x^{(i)}, r^{(i)}, z, c|\Theta)$ denotes the complete likelihood for one observation $< u^{(i)}, e^{(i)}, x^{(i)}, r^{(i)} >$ and values of z and c. The difference $\mathcal{F}(Q,\Theta) - \log p(\mathcal{D}|\Theta)$ is KL-divergence of $Q^{(i)}_{zc}$ with the posterior probability $p(z,c|u^{(i)}, e^{(i)}, x^{(i)}, r^{(i)})$ [30] to minimize. In the E-step of the EM algorithm, we maximize $\mathcal{F}(Q,\Theta)$ w.r.t $Q^{(i)}_{zc}$ to get $\hat{Q}^{(i)}_{zc}$ and in the M-step, the maximization is made w.r.t. Θ to get $\hat{\Theta}$ which is also the local maximum of $\log p(\mathcal{D}|\Theta)$. The optimal distribution $\hat{Q}$ is computed in the E-step as the posterior $p(z,c|u^{(i)}, e^{(i)}, x^{(i)}, r^{(i)})$ of hidden variables given the observed ones

$$\hat{Q}^{(i)}_{zc} = \frac{\hat{\theta}^U_{zu^{(i)}} \hat{\theta}^E_{ze^{(i)}} \hat{\theta}^R_{zcr^{(i)}} \hat{\theta}^D_{cd^{(i)}} p(\vec{v}^{(i)}|c)}{\sum_{z=1}^{K}\sum_{c=1}^{M} \hat{\theta}^U_{zu^{(i)}} \hat{\theta}^E_{ze^{(i)}} \hat{\theta}^R_{zcr^{(i)}} \hat{\theta}^D_{cd^{(i)}} p(\vec{v}^{(i)}|c)}. \quad (9)$$

The hat on probabilities indicates quantities parameterized by the estimated $\hat{\Theta}$. We give the formulas of M-step in the following:

$$\hat{\theta}^Z_z = \frac{\sum_i \sum_c \hat{Q}^{(i)}_{zc}}{N}, \hat{\theta}^C_{zc} = \frac{\sum_i \hat{Q}^{(i)}_{zc}}{\mathrm{N}} \hat{\theta}^Z_z$$

$$\hat{\theta}^D_{cd} = \frac{\sum_{i:d^{(i)}=d} \sum_z \hat{Q}^{(i)}_{zc}}{N \sum_z \hat{\theta}^Z_z \hat{\theta}^C_{zc}}$$

$$\hat{\theta}^U_{zu} = \frac{\sum_{i:u^{(i)}=u} \sum_c \hat{Q}^{(i)}_{zc}}{N \hat{\theta}^Z_z}$$

$$\hat{\theta}^E_{ze} = \frac{\sum_{i:e^{(i)}=e} \sum_c \hat{Q}^{(i)}_{zc}}{N \hat{\theta}^Z_z}$$

$$\hat{\theta}^R_{zcr} = \frac{\sum_{i:r^{(i)}=r} \hat{Q}^{(i)}_{zc}}{N \hat{\theta}^Z_z \hat{\theta}^C_{zc}}.$$

We use the Fisher scoring method to update the parameters of the Dirichlet distributions $p(\vec{v}|c)$ in the M-step by following a similar methodology given by [27]. The Dirichlet parameters $\vec{\alpha}^c$ are initialized using a fuzzy C-Means [27] algorithm. We run the EM algorithm for each initialization and we maintain parameter values that ensure the highest likelihood.

V. Modeling $s(x,u,e|\mathcal{X}^{ue}_t)$

Now we will focus on modeling $s(x,u,e|\mathcal{X}^{ue}_t)$ such that the recommendation list is diversified. The following utility is appropriate for the recommendation [24], [22]:

$$s(x,u,e) = \log \frac{p(r^+|x,u,e)}{p(r^-|x,u,e)} \quad (11)$$

where $p(r^-|x,u,e) = \sum_{r=1}^{T_r} p(r|x,u,e)$, $p(r^+|x,u,e) = 1 - p(r^-|x,u,e)$ and T_r is a threshold used to separate positive and negative ratings. We notice that the quantities $p(r|x,u,e)$ can be derived from (4) by conditioning $p(u,e,x,r)$ on (u,e,x). The threshold T_r can be set as a neutral vote once by an expert for all the users (e.g., $T_r = 3$ for a five star scale [31]). Also, one may use T_r^u, a threshold personalized to each user's definition of relevance as in adaptive filtering [32].

Major works on diversity have been made within IR [23], [21] which employ greedy optimizations. Indeed, the first document is selected as the most similar to the query (topic). After that, documents are inserted successively into the result set according to both their relevance to the query and the redundancy they provide with respect to the already retrieved documents. For example, the authors in [33], compute for each document a score as a weighted sum of the document's redundancy (i.e., maximum similarity with previous documents) and its relevance (i.e., similarity with the query). In [23], the authors address the problem of diversity by penalizing the results with lower number of covered subtopics. The authors in [22], define a new metric k-call at rank n for measuring the minimum number of relevant documents (i.e., k) within a result set of n documents. They employ a unified model where the prior distribution (over word features) is updated successively each time a document is selected within the result set. In [21], the authors define another unified model for adjusting this prior distribution through an "extended" shrinkage smoothing based on a new mixture model. We notice that these unified models cannot be employed directly for CBIS since the "bag-of-words" paradigm is not appropriate for visual features especially with few data (L is small).

From consumer psychology [2], one could affirm that reaching the highest diversity of $\mathcal{X}^{ue}$ can be achieved by maximizing the metric 1-call at rank L proposed in [22]. This metric denotes the probability to have *"at least one"* relevant document within $\mathcal{X}^{ue}$. We follow a similar methodology of [22] which we adapt to CBIS. For a given set $\mathcal{X}_t^{ue}$, we select the new document from a certain subset of $\mathcal{X}$ where "all" documents are "visually" and "semantically" dissimilar from those in $\mathcal{X}_t^{ue}$. To achieve that, we assume that visual documents in $\mathcal{X}_t^{ue}$ are *"irrelevant"* which leads to

$$p\big(r_t^+|u, e, x_t^{ue}, x_1^{ue}, r_1^-, \ldots, x_{t-1}^{ue}, r_{t-1}^-\big) \tag{12}$$

where $r_j^+ (r_j^-)$ denotes the positive (negative) rating associated to $x_j^{ue}, j = 1, \ldots, t$. In order to take into account the new information about the irrelevance of $\mathcal{X}_t^{ue}$, the parameters of the model $\Theta^{(t)}$ need to be updated from the observation $< u, e, x_{t-1}^{ue}, r- >$ each time x_{t-1}^{ue} is recommended. Let $s(x, u, e; \Theta)$ be the utility (11) computed using a certain model Θ given by (6). Then, x_t^{ue} is selected by maximizing $s(x, u, e; \Theta^{(t)})$ with updated parameters as

$$s(x, u, e|\mathcal{X}_t^{ue}) = s(x, u, e; \Theta^{(t)}). \tag{13}$$

Initially, we set $\mathcal{X}_1^{ue} = \emptyset$ and $\Theta^{(1)} = \Theta_{\rm ML}$ obtained after the ML estimation presented in Section IV. From the fact that any observation $< u, e, x_{t-1}^{ue}, r- >$ does not change neither the user classes z nor visual document classes c, then we assume that the only parameters to update are $\vec{\theta}_{zc}^R$. Hence, we consider two documents as "similar" if they belong to the same content class. We use the hard assignment approach [34] by which we update only $\theta_{zc_{t-1}^*}^R$, where c_{t-1}^* is the content class of x_{t-1}^{ue}, using only "one iteration" of the EM algorithm. In E-step, we compute $\hat{Q}_{zc}^{(t-1)}$ for all z, c using (9). The M-step is given by

$$c_{t-1}^* = \arg\max_c \sum_{z=1}^{K} \hat{Q}_{zc}^{(t-1)},$$
$$\theta_{zc_{t-1}^* r^-}^{R\ \ (t)} = \frac{\hat{Q}_{zc_{t-1}^*}^{(t-1)}}{\hat{Q}_{zc_{t-1}^*}^{(t-1)}} = 1. \tag{14}$$

Intuitively, (14) allows the selection of image class representatives with the highest predicted ratings. All the visual documents belonging to an already selected content class are considered irrelevant which seems natural [35]. Therefore, this ranking strategy seeks for "novel" and "most relevant" visual documents. We notice that after having recommended $\mathcal{X}^{ue}$, the parameter set Θ are initialized to $\Theta_{\rm ML}$ for future recommendations. Also, considering $\Theta^{(t)} = \Theta_{\rm ML}, \forall t$, the model (13) turns to a rating-based ranking (i.e., L-call at rank L metric [22]) given by

$$s(x, u, e|\mathcal{X}_t^{ue}) = s(x, u, e; \Theta_{\rm ML}). \tag{15}$$

VI. Computational Complexity

Real recommender systems often manipulate a huge amount of data. Therefore, it would be useful to evaluate the computational complexity of the proposed VCAF model by analyzing the computational effort during both phases of learning and suggestion. We assume a training data set of N observations. In one E-step, MKN variational probabilities $\hat{Q}$ are computed using arithmetic operations except for $p(\vec{v}|c)$ which can be computed only once before each E-step for all c in $O(nM)$ operations (dimension of $\vec{v}$). In practice $n <$ KN and the cost of one E-step is O(MKN). In the M-step, the parameters $\vec{\theta}_\pi^A$ are updated by summing out posterior probabilities for all observations. The biggest effort is deployed in computing the quantities θ_z^Z which require O(KNM) operations. For each Dirichlet parameter $\vec{\alpha}^c$, the Fisher update step requires the computation of the Gradient and the Hessian of the log-likelihood w.r.t $\vec{\alpha}^c$. The cost of computing the Gradient vector is $O(nKN)$. On the other hand, the calculation of the Hessian is proportional to n^2 (nK recalculations are avoided by considering $\vec{\theta}_z^C$). We notice that the Hessian is a matrix with a special structure and its inverse can be computed in $O(n^2)$ operations. By considering M Dirichlet distributions and since $n << KMN$, then the complexity of one EM iteration is $O(nKMN)$. In practice, the EM algorithm requires 30–80 iterations to converge.

For the complexity of the suggestion, we assume a recommendation list $\mathcal{X}^{ue}$ of L visual documents and two ranking strategies: rating-based and diversity-based given by (15) and (13), respectively. Both strategies require computing the utility $s(u, e, x)$ for each user, context and visual document. In rating-based ranking, the cost of computing $s(u, e, x)$ is $O(KM)$. On the other hand, the diversity-based ranking involve updating the parameters of the model and requires $O(KML)$ arithmetic operations.

VII. Experiments: Pilot Study

In this section, we present our method for evaluating the VCAF model. We have made two kinds of evaluations: contextual and noncontextual. The first one focuses on measuring the accuracy of the data modeling provided by the generative model while the second evaluation measures the usefulness of the visual content recommendation in content based image retrieval (CBIR).

A. Data Set

The collection of visual documents we have used in experiments contains 4775 images annotated with 87 discriminative keywords collected in part from Washington University (http://www.cs.washington.edu/research/imagedatabase) and another part from collections of free photographs on the Internet. This collection is diversified and contains both natural (landscapes, sea, mountains, etc.) and man-made images (buses, bridges, etc.). For annotation, both manual and automatic techniques may be adopted. However, manual annotation may be very time-consuming while entirely-automatic annotation may be not efficient. In this work, we opted for a compromise consisting in semi-automatic annotation. First, some images are picked from the collection and annotated manually. The same images are then introduced as queries to a CBIR system [36]. Finally, the retrieved (similar) images inherit the same keywords as the queries. Once images are annotated, we extract keyword features by computing $p(x, w_j)$ the joint probability to observe the keyword w_j associated to the image x [37]. After extraction of these probabilities, we consider a keyword feature as the vector of all these probabilities. For visual content characterization, we have employed both local and global descriptors. For local descriptors, we use the Scale Invariant Feature Transform (SIFT) to represent image patches. This descriptor has been used with success in object recognition and has provided the best performance for matching. We cluster SIFT vectors using K-Means which provides a visual vocabulary as the set of cluster centers or keypoints. After that, we generate for each image a normalized histogram of frequencies of each keypoint ("bag of keypoints") [38]. We have found that a number of 100 keypoints provided a good clustering for our collection. For global descriptors, we used the color correlogram [39] for image texture representation, and the edge histogram descriptor [40]. The color correlogram is built by considering the spatial arrangement of colors in the image for different displacements. Then, we compute for each displacement nine parameters that are the mean, variance, homogeneity, energy, contrast, entropy, correlation, cluster prominence and cluster shade. The edge histogram descriptor represents the frequencies of four edge orientations $(0, \pi/4, \pi/2, \mathrm{and} 3\pi/4)$. We define the context by the location $\mathcal{L} = \mathrm{in-campus}, \mathrm{out-campus}$ and the time $\mathcal{T} = (\mathrm{weekday}, \mathrm{weekend})$. For the rating scale we define it as $\mathcal{R} = 1, 2, \ldots, 5$.

1) User Preferences: Eight human subjects participated in the experiments. We notice that the VCAF model cannot be learned from the data provided by only eight subjects (few data). Therefore, in order to validate our approach, we have built a data set in two steps: the collection of 24 user preferences (UP) and the generation of the data set. First, each subject provided three UP (profiles) he or she considered as the most realistic where one of them is his/her own profile. Indeed, for each context and UP, each subject provided a rating on some content class representatives selected randomly from the image collection (see Fig. 4). In total, for each image, the subject provided 12 ratings since we have four possible contexts and three UP per subject.

Fig. 4. User interface that collects ratings from eight human subjects. If the option "main profile" is activated, then, the subject provides relevance degree according to his or her own profile on an image selected randomly from the collection. The two other profiles correspond to preferences judged "realistic" by the subject.

In the second step, we build a data set $\mathcal{D}$ of ratings for 50 simulated users from the collected 24 UP. For each simulated user u, we associate randomly one UP and we generate 60 ratings per context on images selected randomly from the collection. The rating of each image d, is the value associated to its content-class according to the selected UP and context.

In order to evaluate the sensitivity of the rating prediction according to the number of user preferences, we consider two data sets $\mathcal{D}_9$ and $\mathcal{D}_{24}$ with the same number of users. $\mathcal{D}_9$ denotes the set of ratings sampled from the preferences provided by the first three human subjects while $\mathcal{D}_{24}$ is the set of the ratings sampled from 24 user preferences (i.e., $\mathcal{D}$). We learn and evaluate the VCAF and some other algorithms on each subset separately.

B. Noncontextual Evaluation

The objective of this evaluation is to evaluate the performance of the proposed generative model of (4) in predicting accurate ratings.

1) Experiment Protocol: To show the benefits of using the visual information, we evaluate the following algorithms (variants) of the proposed VCAF model depending on the content description (Keywords K, Visual features V or KV) and the contextual information (context-aware C or not). Indeed, we evaluate VCAF-KV, VCAF-KC, VCAF-VC, VCAF-KVC. We compare our algorithm with the Aspect [16], FMM [17] and Exponential [20] models. The Exponential model (hybrid IF) was proposed for filtering text documents by using normalized term frequencies to describe documents. In this work, we use

TABLE I
AVERAGE MAE OVER 10 RUNS PROVIDED BY THE DIFFERENT ALGORITHMS ON $\mathcal{D}_9$ AND $\mathcal{D}_{24}$ USING 10-FOLD CROSS VALIDATION

data set	$\mathcal{D}_9$		$\mathcal{D}_{24}$	
Ratings	*All*	*Extreme*	*All*	*Extreme*
Aspect	1.262	1.381	1.412	1.552
baseline	*0*	*0*	*0*	*0*
FMM	1.227	1.271	1.315	1.362
(%)	*2.77*	*7.97*	*6.87*	*12.24*
Exponential	1.182	1.199	1.264	1.283
(%)	*6.34*	*13.17*	*10.48*	*17.33*
VCAF-KV	1.178	1.193	1.257	1.272
(%)	*7.13*	*13.61*	*10.98*	*18.04*
FMM-C	0.107	0.118	0.304	0.325
(%)	*91.52*	*91.46*	*78.47*	*79.05*
VCAF-KC	0.088	0.097	0.282	0.301
(%)	*93.03*	*92.48*	*80.03*	*80.61*
VCAF-VC	0.081	0.095	0.215	0.240
(%)	*93.00*	*92.64*	*84.77*	*84.54*
VCAF-KVC	**0.076**	**0.085**	**0.188**	**0.221**
(%)	***94.27***	***93.41***	***86.68***	***85.76***

both keywords and visual features to represent visual documents for the Exponential model. We evaluate also FMM-C by representing each visual document x by a categorical index d in (4).

We use the Mean Absolute Error (MAE) [15] to measure the accuracy of the prediction. We learn the models using a part of the data set (i.e., learning data set) and we measure the average absolute deviation between the actual and the predicted ratings. For VCAF, we predict a rating as $r^* = \arg\max_r p(r|u, e, x)$. We employ the 10-fold cross validation in which we compute an averaged MAE over 10 runs. It has been argued [13] that MAE is more critical for extreme ratings (i.e., $r = 1$ or $r = 5$ in our case) for which recommendation decisions such as discard or recommend items are generally made by most recommenders. For that end, we will measure also MAE for extreme ratings separately as shown in Table I. Finally, we measure the MAE for unrated images whose ratings are predicted based on the content features only as explained in Section III-A. We compute also the improvement (%) of an algorithm with respect to a baseline as: $\% = 100 * (\text{MAE} - \text{MAE}_{\text{baseline}})/\text{MAE}_{\text{baseline}}$.

2) Results: The first four rows of Table I show the added value of content features. Indeed, a Student's t-test shows that both Aspect and FMM are significantly outperformed by VCAF and Exponential models on both data sets $\mathcal{D}_9$ and $\mathcal{D}_{24}$. The last four rows of Table I illustrate the influence of the context on the performance of the algorithms. For example, FMM-C has provided a statistically significant improvement of accuracy 91.28% (76.88%) than FMM on $\mathcal{D}_9$ $(\mathcal{D}_{24})$ according to the Student's t-test. Similarly, VCAF-KVC is significantly better 93.55% (85.04%) than VCAF-KV on $\mathcal{D}_9$ $(\mathcal{D}_{24})$. This fact raised because, in our data sets, some visual documents were preferred within a context while they were disliked in another context by the same user. Thus, in accordance with researches in consumer psychology [4], [3], this experiment demonstrates the importance of the context in modeling user preferences. We notice also that VCAF-KVC has provided the lowest prediction error which outlines that both visual and textual features are two complementary representations that improve the prediction accuracy of existing CF methods. Finally, Table I shows a significant decrease in the performance of all the algorithms on $\mathcal{D}_{24}$ comparatively with $\mathcal{D}_9$. For a fixed number of users, the higher the number of user preferences, the lower the performance of the algorithm.

Fig. 5(a) shows that VCAF maintains similar (at most $+7.69\%$) prediction accuracies for $\simeq 480$ nonrated images on $\mathcal{D}_{24}$. In addition, these errors remain under $< 29\%$ for approximately 1000 nonrated images. Fig. 5(b) shows the influence of the amount of the available knowledge (i.e., data sparsity) about the user on the prediction accuracy. Indeed, when each user provided at least 20 ratings, the figure reports a slight decrease in the performance of all algorithms (at most $\simeq 22\%$) with respect to their performance in Table I (50 ratings/user). Moreover, multidimensional content descriptors fitted using Dirichlet mixtures, allowed the VCAF model to resist to data sparsity and reported a decrease in performance $\simeq 53.81\%$ in the case of 4 ratings/user.

It should be stressed that experiments on simulated data may not provide accurate conclusions about the performance of real systems. Indeed, we have made other experiments on a real data set collected from 27 students in the faculty of science. The significant improvement provided by both VCAF-KV and VCAF-KVC was 25.83% and 37.17%, respectively comparatively with the Aspect model. We notice that this data set remains too small and can not help in assessing the scalability of VCAF on real systems.

C. Contextual Evaluation

This evaluation measures the usefulness of the proposed method in a concrete application that is the content based image retrieval. We evaluate two ranking methods: rating-based of (15) and diversity-based given by (13).

In each experiment run, we initialize $\mathcal{X}^{ue} = \emptyset$ and we put $T_r = 3$ to separate negative and positive ratings. We address the *"page zero problem"* [41] of a CBIR system. Usually, CBIR systems select randomly an initial set (i.e., page zero) of images from which a user would build a search query. We propose to recommend images for the page zero according to the preferences of each user and his/her context instead of the random selection currently provided by CBIR systems.

1) Experiment Protocol: We use the data set $\mathcal{D}$ generated previously and we collect satisfaction indicators from eight human subjects who participated in the generation of that data set. Each subject is recommended eight images in the page zero according to his/her profile. Then, the subject uses each of these images as a query for a CBIR system *AtlasWise* [36] to look for the image corresponding to his/her information need. After that, the subject attributes a binary relevance degree ("0" or "1") to each recommended image in the page zero. To evaluate the usefulness of each algorithm, we use both the precision for a certain scope (i.e., size of the page zero) [10] and the number of query refinements necessary to satisfy the subject's information need. We compute the precision as the percentage of relevant images in the page zero. In order to show the benefits of the recommendation, we compare the VCAF model with the random selection (*Random*). Finally, we study the number of query refinements for different scopes with the diversity-based

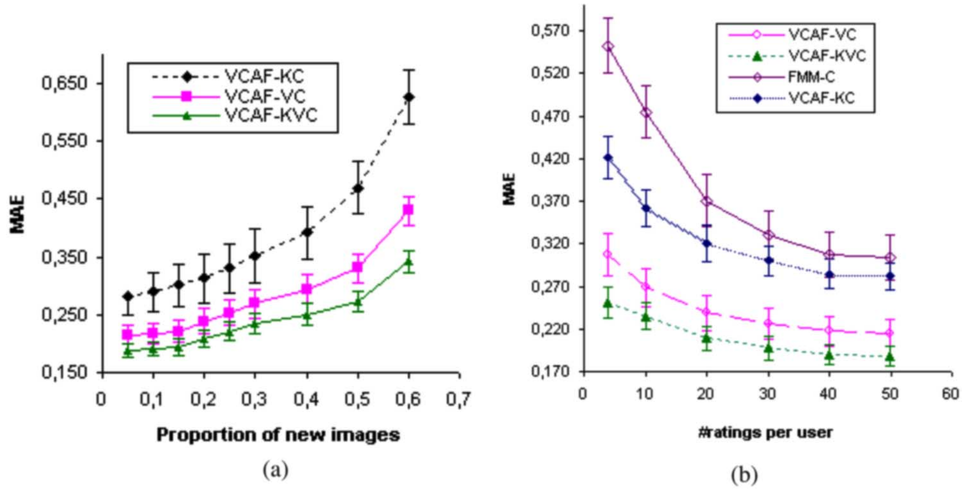


Fig. 5. MAE curves with error bars for different new images ratios and number of ratings per user on the data set $\mathcal{D}_{24}$. (a) New images ratios and (b) MAE for different number of training ratings per user.

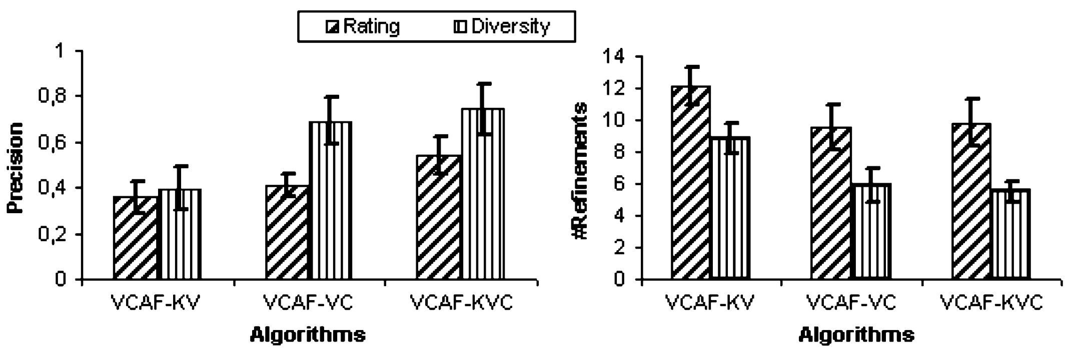


Fig. 6. Average precision and number of refinements with error bars reported by eight subjects for a scope of 8 images.

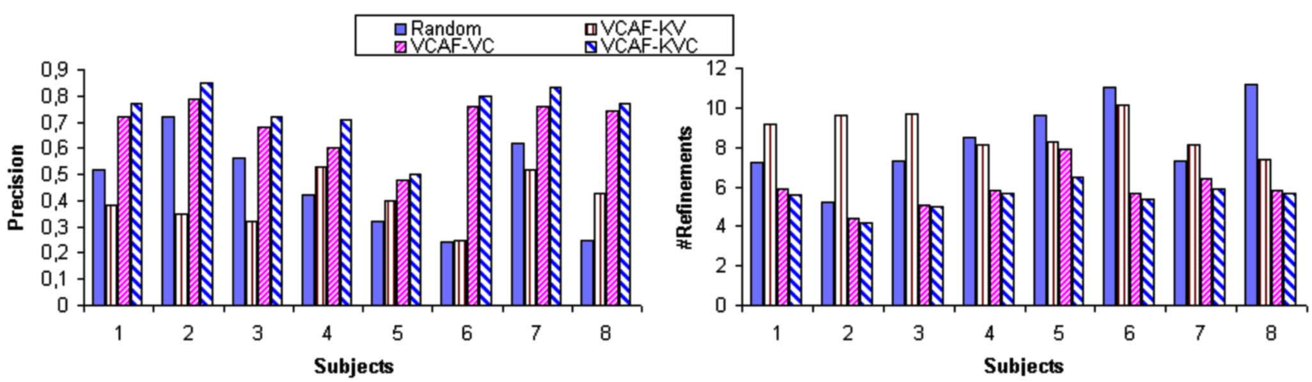


Fig. 7. Average precision and number of refinements reported by VCAF with diversity ranking and Random algorithms. Satisfaction indicators are provided by each subject for a scope of 8 images.

ranking as shown in Fig. 8. We report the average precision and query refinements with scope of eight images for each subject individually in Fig. 7 and global performance over all subjects in Fig. 6.

We can see clearly from Fig. 6 that diversified recommendations have provided lower refinements and higher precision than rating-based ones. For example, VCAF-VC with diversity ranking has improved the precision by 67.07% and the number of refinements by 38.21% comparatively with rating-based VCAF-VC. This fact conforms with researches in consumer psychology [2] and IR [21], [22]. However, a diversified page zero does not necessarily improve the retrieval due to the performance reported for VCAF-KV with lower precision 45.95% and higher number of refinements 60.00% than VCAF-KVC. According the previous experiment in Section VII-B, we explain the reported performance of VCAF-KV by errors in the rating prediction than VCAF-KVC. From this, we notice the importance of both the rating and diversity ranking on the quality of the page zero.

Fig. 7. reports the performance of VCAF with diversity ranking and Random algorithms for each subject separately. We notice that some subjects (1,2,3,7) have reported a better experience for the random selection than VCAF-KV (e.g., subject 2: higher precision 9.67%, improvement of refinements

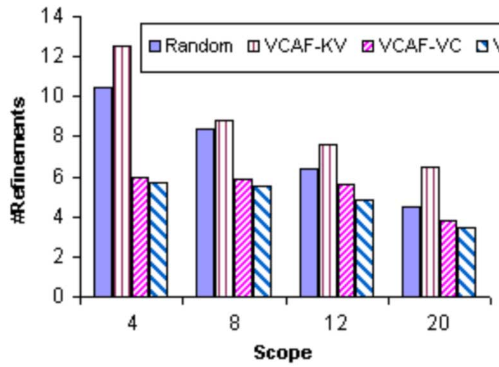


Fig. 8. Average number of refinements for different scope values reported by VCAF with diversity ranking and Random algorithm.

6.85%). The preferences of these subjects are highly influenced by the contextual situation. Therefore, the predicted ratings by VCAF-KV were very erroneous for those subjects which penalized the page zero. On the other hand, the preferences of the subject 5 were "less sensitive" to the context since the number of refinements with VCAF-KV was slightly higher (4.81%) than the one reported with VCAF-KVC. We notice that all the subjects have reported the best appreciation for the page zero provided by VCAF-KVC.

From Fig. 8, we see that the higher the value of the scope, the lower the number of refinements. Also, for a scope of 20, the random selection requires only 28.57% more queries than VCAF-VC. As the size of the page zero increases, the different methods tend to give similar performances. Generally, subjects use one or two images in queries and the availability of many images selected randomly increases the chance to find at least one image relevant.

VIII. Conclusion

Recommender systems were extensively studied in literature and have not addressed the added value of the visual information. In this paper, we proposed content-based image suggestion (CBIS) which exploits both visual and textual information in making useful recommendations. We have studied the influence of the following factors on user preferences: 1) content features (visual and textual), 2) contextual situations, and 3) users' needs for diversified recommendations. The proposed model predicts the outcome of the user's decision making by considering both his/her conformity to the community of users with similar interests and his/her need for variety. We have measured the user's conformity by a generative graphical model which incorporates both content and context information to identify user and visual document classes even with sparse data. Under the non-creation of new content classes, the proposed generative model predicted ratings for new visual documents (up to 30% of the whole collection) with acceptable accuracies. On the other hand, we addressed the user's need for novel and diversified content by using dissimilarity information with his/her already seen visual documents. Our pilot study showed that the accuracy of the rating's prediction is greatly influenced by the content features used to describe images and by the user's contextual situation.

Acknowledgment

The authors would like to thank the associate editor and anonymous referees for their helpful comments. For example, the recommendation utility in (11) was suggested by an anonymous referee. The authors also thank also Prof. Nizar Bouguila.

References

[1] P. Messaris, *Visual Persuasion: The Role of Images in Advertising*. Thousand Oaks, CA: Sage, 1997.

[2] E. Hirschman, "Innovativeness, novelty seeking, and consumer creativity," *J. Consum. Res.*, vol. 7, no. 3, pp. 283–295, Dec. 1980.

[3] T. R. Graeff, "Consumption situations and the effects of brand image on consumers brand evaluations," *Psychol. Marketing*, vol. 14, no. 1, pp. 49–70, 1997.

[4] R. W. Belk, "Situational variables and consumer behavior," *J. Consumer Res.*, vol. 2, no. 3, pp. 157–164, 1975.

[5] B. E. Kahn, "Consumer variety-seeking among goods and services—an integrative review," *J. Retail. Consumer Serv.*, vol. 2, no. 3, pp. 139–148, 1995.

[6] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, Jun. 2005.

[7] G. J. F. Jones and P. Brown, "Challenges and opportunities of context-aware information access," in *Proc. Int. Workshop Ubiquitous Data Manage.*, 2005, pp. 53–62.

[8] C. Y. Kim, J. K. Lee, Y. H. Cho, and D. H. Kim, "VISCORS: A visual-content recommender for the mobile web," *IEEE Intell. Syst.*, vol. 19, no. 6, pp. 32–39, 2004.

[9] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "Grouplens: An open architecture for collaborative filtering of netnews," in *Proc. ACM Conf. Comput. Supported Cooperative Work*, 1994, pp. 175–186.

[10] M. L. Kherfi, D. Ziou, and A. Bernardi, "Image retrieval from the world wide web: Issues, techniques, and systems," *ACM Comput. Surveys*, vol. 36, no. 1, pp. 35–67, 2004.

[11] D. W. Oard, "The state of the art in text filtering," *User Modeling User-Adapted Interaction*, vol. 7, no. 3, pp. 141–178, 1997.

[12] R. Mooney and L. Roy, "Content-based book recommending using learning for text categorization," in *Proc. 5th ACM Conf. Digital Libraries*, 2000, pp. 195–204.

[13] K. Yu, A. Schwaighofer, V. Tresp, X. Xu, and H.-P. Kriegel, "Probabilistic memory-based collaborative filtering," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 1, pp. 56–69, Jan. 2004.

[14] R. Jin, L. Si, and C. Zhai, "A study of mixture models for collaborative filtering," *J. Inf. Retrieval*, vol. 9, no. 3, pp. 357–382, 2006.

[15] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," in *Proc. 14th Conf. Uncertainty Artificial Intell.*, 1998, pp. 43–52.

[16] T. Hofmann, "Latent semantic models for collaborative filtering," *ACM Trans. Inf. Syst.*, vol. 22, no. 1, pp. 89–115, Jan. 1, 2004.

[17] L. Si and R. Jin, "Flexible mixture model for collaborative filtering," in *Proc. 20th Int. Conf. Mach. Learning*, 2003, pp. 704–711.

[18] A. Popescul, L. H. Ungar, D. M. Pennock, and S. Lawrence, "Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments," in *Proc. 17th Conf. Uncertainty Artificial Intell.*, 2001, pp. 437–444.

[19] P. Melville, R. J. Mooney, and R. Nagarajan, "Content-boosted collaborative filtering for improved recommendations," in *Proc. 18th Nat. Conf. Artificial Intell.*, 2002, pp. 187–192.

[20] L. Si and R. Jin, "Unified filtering by combining collaborative filtering and content-based filtering via mixture model and exponential model," in *Proc. 13th ACM Int. Conf. Inf. Knowledge Manage.*, 2004, pp. 156–157.

[21] Y. Zhang, J. Callan, and T. Minka, "Novelty and redundancy detection in adaptive filtering," in *Proc. 25th Int. ACM Conf. Res. Dev. Inf. Retrieval*, 2002, pp. 81–88.

[22] H. Chen and D. Karger, "Less is more: Probabilistic models for retrieving fewer relevant documents," in *Proc. 29th Int. ACM Conf. Res. Dev. Inf. Retrieval*, 2006.

[23] C. Zhai, W. W. Cohen, and J. Lafferty, "Beyond independent relevance: Methods and evaluation metrics for subtopic retrieval," in *Proc. 26th Int. ACM Conf. Res. Dev. Inf. Retrieval*, 2003, pp. 10–17.

[24] K. W. Cheung, K. C. Tsui, and J. Liu, “Extended latent class models for collaborative recommendation,” *IEEE Trans. Syst. Man Cybern. A*, vol. 34, no. 1, pp. 143–148, Jan. 2004.

[25] M. Datcu, H. Daschiel, A. Pelizzari, M. Quartulli, A. Galoppo, A. Colapicchioni, M. Pastori, K. Seidel, P. G. Marchetti, and S. D’Elia, “Information mining in remote sensing image archives: System concepts,” *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 12, pp. 2923–2936, Dec. 2003.

[26] J. Pearl, *Causality : Models, Reasoning, and Inference*. New York: Cambridge Univ. Press, 2000.

[27] N. Bouguila, D. Ziou, and J. Vaillancourt, “Unsupervised learning of a finite mixture model based on the Dirichlet distribution and its application,” *IEEE Trans. Image Process.*, vol. 13, no. 11, pp. 1533–1543, Nov. 2004.

[28] N. Bouguila, D. Ziou, and J. Vaillancourt, “Novel mixtures based on the Dirichlet distribution: Application to data and image classification,” in *Proc. Mach. Learning Data Min. Pattern Recognition*, Leipzig, Germany, Jul. 2003, pp. 172–181, LNCS 2734.

[29] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions.*. New York: Wiley, 1997.

[30] M. I. Jordan, Z. Ghahramani, T. Jaakkola, and L. K. Saul, “An introduction to variational methods for graphical models,” *Mach. Learn.*, vol. 37, no. 2, pp. 183–233, 1999.

[31] D. Billsus and M. J. Pazzani, “Learning collaborative information filters,” in *Proc. 15th Int. Conf. Mach. Learning*, 1998, pp. 46–54.

[32] S. Robertson and S. Walker, “Threshold setting in adaptive filtering,” *J. Document.*, vol. 56, no. 3, pp. 312–331, May 1, 2000.

[33] J. Carbonell and J. Goldstein, “The use of MMR, diversity-based reranking for reordering documents and producing summaries,” in *Proc. 21st Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retrieval*, 1998, pp. 335–336.

[34] J. Zhang, Z. Ghahramani, and Y. Yang, “A probabilistic model for online document clustering with application to novelty detection,” in *Proc. Adv. Neural Inf. Process. Syst. 17*, 2004, pp. 1617–1624.

[35] W. Dai and R. K. Srihari, “Minimal document set retrieval,” in *Proc. 14th ACM Int. Conf. Inf. Knowledge Manage.*, 2005, pp. 752–759.

[36] M. L. Kherfi and D. Ziou, “Relevance feedback for CBIR: A new approach based on probabilistic feature weighting with positive and negative examples,” *IEEE Trans. Image Process.*, vol. 15, no. 4, pp. 1017–1030, Apr. 2006.

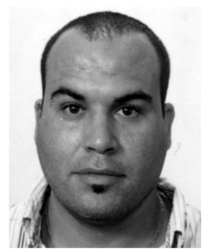
[37] M. L. Kherfi, D. Brahmi, and D. Ziou, “Combining visual features with semantics for a more effective image retrieval,” in *Proc. 17th Int. Conf. Pattern Recognition*, 2004, pp. 961–964.

[38] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, “Visual categorization with bags of keypoints,” in *Proc. ECCV Workshop Stat. Learning Comput. Vision*, 2004, pp. 1–22.

[39] J. Huang, S. R. Kumar, M. Mitra, W. J. Zhu, and R. Zabih, “Image indexing using color correlograms,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognition*, San Juan, Puerto Rico, 1997, pp. 762–768.

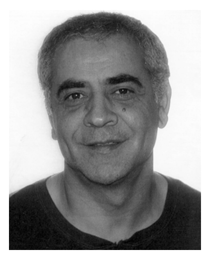
[40] A. K. Jain and A. Vailaya, “Image retrieval using color and shape,” *Pattern Recognit.*, vol. 29, no. 8, pp. 1233–1244, Aug. 1996.

[41] M. L. Cascia, S. Sethi, and S. Sclaroff, “Combining textual and visual cues for content-based image retrieval on the world wide web,” in *Proc. IEEE Workshop Content-Based Access Image Video Libraries*, 1998.



Sabri Boutemedjet received the B.Eng. degree in computer science from the University of Annaba, Algeria, in 1999 and the M.Sc. degree in computer science from the Université du Québec á Montréal (UQAM), Montreal, QC, Canada, in 2004. He is currently pursuing the Ph.D. degree at the Université de Sherbrooke, Sherbrooke, QC, Canada.

His research interests include information filtering and retrieval, data mining, and pattern recognition.



Djemel Ziou received the B.Eng. degree in computer science from the University of Annaba, Algeria, in 1984 and the Ph.D. degree in computer science from the Institut National Polytechnique de Lorraine (INPL), France, in 1991.

From 1987 to 1993, he was a Lecturer with several universities in France. During the same period, he was a Researcher at the Centre de Recherche en Informatique de Nancy (CRIN) and the Institut National de Recherche en Informatique et Automatique (INRIA) in France. Presently, he is a full Professor in the Department of Computer Science, Universite de Sherbrooke, Sherbrooke, QC, Canada. His research interests include image processing, information retrieval, computer vision, and pattern recognition.

Dr. Ziou is the holder of the Natural Sciences and Engineering Research Council (NSERC)/Bell Canada Research chair in personal imaging. He has served on numerous conference committees as a member or chair. He heads the laboratory MOIVRE and the consortium CoRIMedia, which he founded.