A Graphical Model for Content Based Image Suggestion and Feature Selection

Sabri Boutemedjet[1], Djemel Ziou[1], and Nizar Bouguila[2]

[1] Département d'informatique
Université de Sherbrooke, QC, Canada J1K 2R1
{sabri.boutemedjet,djemel.ziou}@usherbrooke.ca
[2] Concordia Institute for Information Systems Engineering
Concordia University, Montreal, QC, Canada H3G 1T7
bouguila@ciise.concordia.ca

Abstract. Content based image retrieval systems provide techniques for representing, indexing and searching images. They address only the user's short term needs expressed as queries. From the importance of the visual information in many applications such as advertisements and security, we motivate in this paper, the *Content Based Image Suggestion*. It targets the user's long term needs as a recommendation of products based on the user preferences in different situations, and on the visual content of images. We propose a generative model in which the visual features and users are clustered into separate classes. We identify the number of both user and image classes with the simultaneous selection of relevant visual features. The goal is to ensure an accurate prediction of ratings for multidimensional images. This model is learned using the minimum message length approach. Experiments with an image collection showed the merits of our approach.

1 Introduction

Information retrieval (IR) provides tools and techniques that help users to access, browse and summarize information stores 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. The kind of information needs addressed in CBIR is short term. There is another kind of interests i.e. long term or permanent such as desires, tastes and preferences of each user. In today's e-market, products are described using both visual and textual information. From consumer psychology, the visual information has been recognized as an important factor that influences the consumer's decision making and has an important power of persuasion [18]. Indeed, images can convey meanings that cannot be expressed using words. Furthermore, it is well recognized [1] that the consumer choice is also influenced by the external environment or consumer's context defined by the time and location. For example, a consumer could express an information need during a travel that is different from the situation when she or he is working or even at home.

J.N. Kok et al. (Eds.): PKDD 2007, LNAI 4702, pp. 30–41, 2007.

In literature, user preferences are modeled within collaborative filtering (CF) and content based filtering (CBF) communities. CF approaches predict the relevance of a given product for the active user based on the preferences provided by a set of "like-minded" (similar tastes) users. Within the CF framework, each product is represented by its index considered as a categorical variable. The Aspect model [10] and the flexible mixture model (FMM) [24] are examples of some model-based CF approaches which involve the clustering as an underlying principle. The "correct" model order (number of parameters or clusters) was generally chosen "empirically" as a compromise between the model's complexity and the accuracy of recommendation. To the best of our knowledge, the issue of formally identifying the model order from the statistical properties of the data, was not addressed in CF literature. On the other hand, CBF approaches [19] [16], represent the user's profile using content descriptors and infer the relevance of unseen products based on the history of the active user. CBF approaches have targeted mainly textual data such as Web sites and newspapers [16]. Some hybrid approaches that combine CF and CBF [22] have been also proposed taking advantage of both methods.

In this paper, we motivate the *"Content Based Image Suggestion"* (CBIS). CBIS aims at the suggestion of products whose relevance is inferred from the history of users in different contexts on images of the previously consumed products. We try to make a direct "mapping" between products and their visual information described in terms of visual features and/or keywords extracted from images. In this work, we consider an image as a D-dimensional vector $\boldsymbol{v} = (v_1, v_2, \ldots, v_D)$. The visual features may be local such as interest points [15] or global such as color, texture, or shape. The keywords can be automatically or semiautomatically extracted by annotation or recognition process. Therefore, text-based recommendations can be improved by capturing user preferences related to the added-value visual appearance of products. For example, figure 1 shows the list of products preferred by two users. Following a similar methodology in hybrid filtering approaches of text documents [22], the CBIS would consider the two users as "like-minded" since visually, they have preferred the same category of products ("motorbikes"). Then, the "camera" can be recommended to the user 2.

In order to predict the relevance of products for users in different contexts, we propose a probabilistic model which we call Visual Content Context-aware Flexible Mixture Model (VCC-FMM). In this model, users and visual documents are clustered separately into homogeneous groups as in FMM [24] except that images are grouped based on an additional visual information.The high dimensionality of visual documents $\boldsymbol{v}$ does not mean necessarily that the clustering structure is contained within the whole set of visual features. Indeed, it is common that high dimensional documents can be clustered based on an unknown set of few features. Moreover, the presence of many irrelevant features may deteriorate the performance of data modeling and increases the computational complexity [20]. The VCC-FMM defines the relevance of each visual feature as the degree of its dependence on class labels [26][21]. In literature [9], the process of feature selection in

mixture models have not received as much attention as in supervised learning. The main reason is the absence of class labels that may guide the selection process in addition to the influence of the considered feature subset on the model order [7]. To address these issues, the VCC-FMM is learned from unlabeled data by minimizing a Minimum Message Length (MML) objective [27].

This paper is organized as follows. The next Section details the VCC-FMM model with an integrated feature selection. In Section 3, we discuss the identification of the model order using MML. Experimental results are presented in Section 4. Finally, we conclude this paper by a summary of the work.

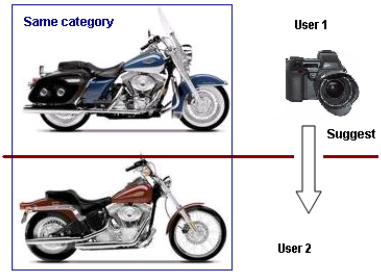


Fig. 1. The principle of Content Based Image Suggestion

2 The Visual Content Context Flexible Mixture Model

We consider a set of users $\mathcal{U} = \{1, 2, \ldots, N_u\}$, a set of visual documents $\mathcal{V} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{N_v}\}$, and a set of possible contexts $\mathcal{E} = \{1, 2, \ldots, N_e\}$. We assume that a numeric rating r measures the relevance of a visual document for a given user and context. This rating r is defined on an ordered voting scale $\mathcal{R} = \{1, 2, \ldots, N_r\}$. First, we model the joint event $p(\boldsymbol{v}, r, u, e)$ by equation (1) where two latent variables c and z label each observation $< u, e, \boldsymbol{v}, r >$. The variable c carries the information about the visual similarity between images while the variable z denote "like-mindedness" of user preferences. The rating r for a given user u, context e and a visual document $\boldsymbol{v}$ can be predicted on the basis of probabilities $p(r|u, e, v)$ that can be derived by conditioning $p(u, e, v, r)$. The conditional independence assumptions among variables are illustrated by the graphical representation of the model in figure 2. The nodes denote random variables and edges (absence) denote conditional dependencies (independencies).

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

where K and M denote the numbers of user classes and image classes, respectively. The quantities $p(z)$ and $p(c)$ denote the a priori weights of user and image classes. $p(u|z)$ and $p(e|z)$ denote the likelihood of a user and context to belong respectively to the user's class z. $p(r|z, c)$ is the probability to generate a rating

for given user and image classes. We model $p(\boldsymbol{v}|c)$ using the Generalized Dirichlet distribution (GDD) [3][2] which is suitable for non Gaussian data such as images. This distribution has a more general covariance structure and provides multiple shapes. The distribution of the c-th component $\boldsymbol{\Theta}_c^*$ is given by:

$$p(\boldsymbol{v}|\boldsymbol{\Theta}_c^*) = \prod_{l=1}^{D} \frac{\Gamma(\alpha_{cl}^* + \beta_{cl}^*)}{\Gamma(\alpha_{cl}^*)\Gamma(\beta_{cl}^*)} v_l^{\alpha_{cl}^* - 1} (1 - \sum_{k=1}^{l} v_k)^{\gamma_{cl}^*} \tag{2}$$

where $\sum_{l=l}^{D} v_l < 1$ and $0 < v_l < 1$ for $l = 1, \ldots, D$. $\gamma_{cl}^* = \beta_{cl}^* - \alpha_{cl+1}^* - \beta_{cl+1}^*$ for $l = 1, \ldots, D-1$ and $\gamma_D^* = \beta_D^* - 1$. In equation (2) we have set $\boldsymbol{\Theta}_c^* = (\alpha_{c1}^*, \beta_{c1}^*, \ldots, \alpha_{cD}^*, \beta_{cD}^*)$. From the mathematical properties of the GDD, we can transform using a geometric transformation a data point $\boldsymbol{v}$ into another data point $\boldsymbol{x} = (x_1, \ldots, x_D)$ with independent features without loss of information [5][2]. In addition, each x_l of $\boldsymbol{x}$ generated by the c-th component, follows a Beta distribution $p_b(.|\theta_{cl}^*)$ with parameters $\theta_{cl}^* = (\alpha_{cl}^*, \beta_{cl}^*)$ which leads to the fact $p(\boldsymbol{x}|\boldsymbol{\Theta}_c^*) = \prod_{l=1}^{D} p_b(x_l|\theta_{cl}^*)$. Therefore, the estimation of the distribution of a D-dimensional GDD sample is indeed reduced to D-estimations of one-dimensional Beta distributions which is very interesting for multidimensional data sets. Since x_l are independent, we can extract *"relevant"* features in the representation space $\mathcal{X}$ as those that depend on class labels [26][21]. In other words, an irrelevant feature is independent of components θ_{cl}^* and follows another background distribution $p_b(.|\xi_l)$ common to all components. Let $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_D)$ be a set of missing binary variables denoting the relevance of all features. ϕ_l is set to 1 when the l-th feature is relevant and 0 otherwise. The "ideal" Beta distribution θ_{cl}^* can be approximated as:

$$p(x_l|\theta_{cl}^*, \phi_l) \simeq \big(p_b(x_l|\theta_{cl})\big)^{\phi_l} \big(p_b(x_l|\xi_l)\big)^{1-\phi_l} \tag{3}$$

By considering each ϕ_l as Bernoulli variable with parameters $p(\phi_l = 1) = \epsilon_{l_1}$ and $p(\phi_l = 0) = \epsilon_{l_2}$ $(\epsilon_{l_1} + \epsilon_{l_2} = 1)$ then, the distribution $p(x_l|\theta_{cl}^*)$ can be obtained after marginalizing over ϕ_l [14] as: $p(x_l|\theta_{cl}^*) \simeq \epsilon_{l_1} p_b(x_l|\theta_{cl}) + \epsilon_{l_2} p_b(x_l|\xi_l)$. The VCC-FMM model is given by equation (4). We notice that the work of [4] is special case of VCC-FMM.

$$p(\boldsymbol{x}, r, u, e) = \sum_{z=1}^{K} \sum_{c=1}^{M} p(z)p(u|z)p(e|z)p(c)p(r|z,c) \prod_{l=1}^{D} [\epsilon_{l_1} p_b(x_l|\theta_{cl}) + \epsilon_{l_2} p_b(x_l|\xi_l)] \tag{4}$$

3 Model Selection and Parameter Estimation Using MML

The variables U, E, R, Φ_l, Z and C are discrete and their distributions are assumed multinomial. We employ the following notation to simplify the presentation. The parameter vector of the multinomial distribution of a discrete variable A conditioned on its parent Π (predecessor) is denoted by $\boldsymbol{\theta}_\pi^A$ (i.e.

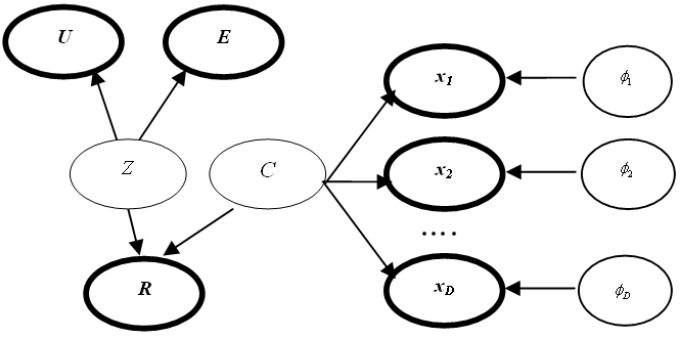


Fig. 2. Graphical representation of VCC-FMM

$A|_{\Pi=\pi} \sim Multi(1;\boldsymbol{\theta}_{\pi}^{A}))$ where $\theta_{\pi a}^{A} = p(A = a|\Pi = \pi)$ and $\sum_a \theta_{\pi a}^{A} = 1$. We have to estimate Θ defined by the parameters of multinomial distributions $\boldsymbol{\theta}_z^U, \boldsymbol{\theta}_z^E, \boldsymbol{\theta}^Z, \boldsymbol{\theta}^C, \boldsymbol{\theta}_z^R, \boldsymbol{\theta}_c^R, \boldsymbol{\theta}^{\phi_l}$ and the parameters of Beta distributions θ_{cl}, ξ_l. We employ the superscripts θ and ξ to denote the parameters of relevant and irrelevant Beta components, respectively (i.e. $\theta_{cl} = (\alpha_{cl}^{\theta}, \beta_{cl}^{\theta})$ and $\xi_l = (\alpha_l^{\xi}, \beta_l^{\xi})$). The log-likelihood of a data set of N independent and identically distributed observations $\mathcal{D} = \{< u^{(i)}, e^{(i)}, \boldsymbol{x}^{(i)}, r^{(i)} > |i = 1, \ldots, N, u^{(i)} \in \mathcal{U}, e^{(i)} \in \mathcal{E}, \boldsymbol{x}^{(i)} \in \mathcal{X}, r^{(i)} \in \mathcal{R}\}$ is given by:

$$\log p(\mathcal{D}|\Theta) = \sum_{i=1}^{N} \log \sum_{z=1}^{K} \sum_{c=1}^{M} p(z)p(c)p(u^{(i)}|z)p(e^{(i)}|z)p(r^{(i)}|z,c)\times \prod_{l=1}^{D}[\epsilon_{l_1} p_b(x_l^{(i)}|\theta_{cl}) + \epsilon_{l_2} p_b(x_l^{(i)}|\xi_l)] \tag{5}$$

The standard Expectation-Maximization (EM) algorithm for maximum likelihood estimation requires a good initialization and the knowledge of both M and K to converge to a good local optimum. Since both M and K are unknown, we employ the MML approach [27] for both estimation of the parameters and identification of K and M. In MML, a penalty term is introduced to the objective of $\mathcal{D}$ to penalize complex models as:

$$MML(K,M) = -\log p(\Theta) + \frac{1}{2}\log|I(\Theta)| + \frac{s}{2}(1 + \log\frac{1}{12}) - \log p(\mathcal{D}|\Theta) \tag{6}$$

where $|I(\Theta)|$, $p(\Theta)$, and s denote the Fisher information, prior distribution and the total number of parameters, respectively. The Fisher information of a parameter is the expectation of the second derivatives with respect to the parameter of the minus log-likelihood. It is common sense to assume an independence among the different groups of parameters which eases the computation of $|I(\Theta)|$ and $p(\Theta)$. Therefore, the joint prior is given by:

$$p(\Theta) = p(\boldsymbol{\theta}^Z)p(\boldsymbol{\theta}^C)[\prod_{z=1}^{K} p(\boldsymbol{\theta}_z^U)p(\boldsymbol{\theta}_z^E)p(\boldsymbol{\theta}_z^R)]\left(\prod_{l=1}^{D}[p(\xi_l)p(\boldsymbol{\epsilon}_l)\prod_{c=1}^{M} p(\theta_{cl})]\right)\prod_{c=1}^{M} p(\boldsymbol{\theta}_c^R) \tag{7}$$

Besides, the Fisher information matrix is bloc-diagonal [8] which leads to $|I(\Theta)| = |I(\boldsymbol{\theta}^Z)||I(\boldsymbol{\theta}^C)| \prod_{c=1}^{M} |I(\boldsymbol{\theta}_c^R)| \Big(\prod_{z=1}^{K} |I(\boldsymbol{\theta}_z^U)||I(\boldsymbol{\theta}_z^E)||I(\boldsymbol{\theta}_z^R)| \Big) \Big(\prod_{l=1}^{D} |I(\xi_l)| |I(\boldsymbol{\epsilon}_l)| \prod_{c=1}^{M} |I(\theta_{cl})| \Big)$. We approximate the Fisher information of VCC-FMM from the complete likelihood which assumes the knowledge of z and c associated to each observation $< u^{(i)}, e^{(i)}, \boldsymbol{x}^{(i)}, r^{(i)} > \in \mathcal{D}$. The Fisher information of the parameters of multinomial distributions can be computed using the result found in [13]. Indeed, if the discrete variable A conditioned on its parent Π, has N_A different values $\{1, 2, \ldots, N_A\}$ in a data set of N observations, then $|I(\boldsymbol{\theta}_\pi^A)| = [\big(Np(\pi)\big)^{N_A - 1}] / [\prod_{a=1}^{N_A} \theta_{\pi a}^A]$, where $p(\pi)$ is the marginal probability of the parent Π. The proposed configuration of VCC-FMM does not involve variable ancestors (parents of parents). Therefore, the marginal probabilities $p(\pi)$ are simply the parameters of the multinomial distribution of the parent variable. Thus,

$$
\begin{aligned}
&|I(\boldsymbol{\theta}_z^R)| = \frac{(N\theta_z^Z)^{N_r - 1}}{\prod_{r=1}^{N_r} \theta_{zr}^R}, \quad |I(\boldsymbol{\theta}_c^R)| = \frac{(N\theta_c^C)^{N_r - 1}}{\prod_{r=1}^{N_r} \theta_{cr}^R}, \quad |I(\boldsymbol{\theta}^Z)| = \frac{N^{K-1}}{\prod_{z=1}^{K} \theta_z^Z} \\
&|I(\boldsymbol{\theta}^C)| = \frac{N^{M-1}}{\prod_{c=1}^{M} \theta_c^C}, \quad |I(\boldsymbol{\theta}_z^U)| = \frac{(N\theta_z^Z)^{N_u - 1}}{\prod_{u=1}^{N_u} \theta_{zu}^U} \\
&|I(\boldsymbol{\theta}_z^E)| = \frac{(N\theta_z^Z)^{N_e - 1}}{\prod_{e=1}^{N_e} \theta_{ze}^E}, \quad |I(\boldsymbol{\theta}^\phi)| = N(\epsilon_{l_1}\epsilon_{l_2})^{-1}
\end{aligned}
\tag{8}
$$

The Fisher information of θ_{cl} and ξ_l can be computed by considering the log-likelihood of each feature taken separately [3]. After the second order derivations of this log-likelihood, we obtain:

$$
\begin{aligned}
|I(\theta_{cl})| &= (N\theta_c^C \epsilon_{l_1})^2 \left| \Big(\psi'(\alpha_{cl}^\theta)\psi'(\beta_{cl}^\theta) - \psi'(\alpha_{cl}^\theta + \beta_{cl}^\theta)(\psi'(\alpha_{cl}^\theta) + \psi'(\beta_{cl}^\theta)) \Big) \right| \\
|I(\xi_l)| &= (N\epsilon_{l_2})^2 \left| \Big(\psi'(\alpha_l^\xi)\psi'(\beta_l^\xi) - \psi'(\alpha_l^\xi + \beta_l^\xi)(\psi'(\alpha_l^\xi) + \psi'(\beta_l^\xi)) \Big) \right|
\end{aligned}
\tag{9}
$$

where Ψ is the trigamma function. In the absence of any prior knowledge on the parameters, we use the Jeffrey's prior for different groups of parameters as the square root of their Fisher information e.g. $p(\boldsymbol{\theta}^Z) \propto \prod_{z=1}^{K} (\theta_z^Z)^{-1/2}$. Replacing $p(\Theta)$ and $I(\Theta)$ in (6), and after discarding the first order terms, the MML objective of a data set $\mathcal{D}$ controlled by VCC-FMM is given by:

$$
\begin{aligned}
MML(K, M) = &\tfrac{N_p}{2} \log N + M \textstyle\sum_{l=1}^{D} \log \epsilon_{l_1} + \sum_{l=1}^{D} \log \epsilon_{l_2} + \tfrac{1}{2} N_p^Z \sum_{z=1}^{K} \log \theta_z^Z \\
&+ \tfrac{1}{2}(N_r - 1) \textstyle\sum_{c=1}^{M} \log \theta_c^C - \log p(\mathcal{D}|\Theta)
\end{aligned}
\tag{10}
$$

with $N_p = 2D(M+1) + K(N_u + N_e + N_r - 1) + MN_r$ and $N_p^Z = N_r + N_u + N_e - 3$. For fixed values of K, M and D, the minimization of the MML objective with

respect to Θ is equivalent to a maximum a posteriori (MAP) estimate with the following improper Dirichlet priors [14]:

$$p(\boldsymbol{\theta}^C) \propto \prod_{c=1}^{M} (\theta_c^C)^{-\frac{N_r-1}{2}}, \quad p(\boldsymbol{\theta}^Z) \propto \prod_{z=1}^{K} (\theta_z^Z)^{-\frac{N_p^Z}{2}}, \quad p(\epsilon_1, \ldots, \epsilon_D) \propto \prod_{l=1}^{D} \epsilon_{l_1}^{-M} \epsilon_{l_2}^{-1} \tag{11}$$

3.1 Estimation of Parameters

We optimize the MML of the data set using the EM algorithm in order to estimate the parameters. The EM algorithm alternates between two steps. In the E-step, the joint posterior probabilities of the latent variables given the observations are computed as:

$$a_{lzc}^{(i)} = p(\phi_l = 1, u^{(i)}, e^{(i)}, x_l^{(i)}, r^{(i)} | z, c, \hat{\Theta}) = \hat{\theta}_{zu^{(i)}}^U \hat{\theta}_{ze^{(i)}}^E \hat{\theta}_{zr^{(i)}}^R \hat{\theta}_{cr^{(i)}}^R \epsilon_{l_1} p(x_l^{(i)} | \hat{\theta}_{cl})$$
$$b_{lzc}^{(i)} = p(\phi_l = 0, u^{(i)}, e^{(i)}, x_l^{(i)}, r^{(i)} | z, c, \hat{\Theta}) = \hat{\theta}_{zu^{(i)}}^U \hat{\theta}_{ze^{(i)}}^E \hat{\theta}_{zr^{(i)}}^R \hat{\theta}_{cr^{(i)}}^R \epsilon_{l_2} p(x_l^{(i)} | \hat{\xi}_l)$$
$$Q_{zci} = p(z, c | u^{(i)}, e^{(i)}, \boldsymbol{x}^{(i)}, r^{(i)}, \hat{\Theta}) = \frac{\hat{\theta}_z^Z \hat{\theta}_c^C \prod_l (a_{lzc}^{(i)} + b_{lzc}^{(i)})}{\sum_{z,c} \hat{\theta}_z^Z \hat{\theta}_c^C \prod_l (a_{lzc}^{(i)} + b_{lzc}^{(i)})} \tag{12}$$

In the M-step, the parameters are updated using the following equations:

$$\hat{\theta}_z^Z = \frac{\max\left(\sum_i \sum_c Q_{zci} - \frac{N_p^Z}{2}, 0\right)}{\sum_z \max\left(\sum_i \sum_c Q_{zci} - \frac{N_p^Z}{2}, 0\right)}, \quad \hat{\theta}_c^C = \frac{\max\left(\sum_i \sum_z Q_{zci} - \frac{N_r-1}{2}, 0\right)}{\sum_c \max\left(\sum_i \sum_z Q_{zci} - \frac{N_r-1}{2}, 0\right)} \tag{13}$$

$$\hat{\theta}_{zu}^U = \frac{\sum_{i:u^{(i)}=u} \sum_c Q_{zci}}{N \hat{\theta}_z^Z}, \quad \hat{\theta_{ze}^E} = \frac{\sum_{i:e^{(i)}=e} \sum_c Q_{zci}}{N \hat{\theta}_z^Z} \quad \hat{\theta}_{cr}^R = \frac{\sum_{i:r^{(i)}=r} \sum_z Q_{zci}}{N \hat{\theta}_c^C} \tag{14}$$

$$\hat{\theta_{zr}^R} = \frac{\sum_{i:r^{(i)}=r} \sum_c Q_{zci}}{N \hat{\theta}_z^Z} \quad \frac{1}{\epsilon_{l_1}} = 1 + \frac{\max\left(\sum_{z,c,i} \frac{Q_{zci} \epsilon_{l_2} p_b(x_l^{(i)} | \xi_l)}{\epsilon_{l_1} p_b(x_l^{(i)} | \theta_{cl}) + \epsilon_{l_2} p_b(x_l^{(i)} | \xi_l)} - 1, 0\right)}{\max\left(\sum_{z,c,i} \frac{Q_{zci} \epsilon_{l_1} p_b(X_{il} | \theta_{cl})}{\epsilon_{l_1} p_b(x_l^{(i)} | \theta_{cl}) + \epsilon_{l_2} p_b(x_l^{(i)} | \xi_l)} - M, 0\right)} \tag{15}$$

The parameters of Beta distributions θ_{cl} and ξ_l are updated using the Fisher scoring method based on the first and second order derivatives of the MML objective [3]. In order to avoid unfavorable local optimums, we use the deterministic EM annealing [25].

The update formulas of θ_c^C, θ_z^Z and ϵ_{l_1} involve a pruning behavior of components and features by forcing some weights to go to zero. It should be stressed that the speed of component pruning for θ_c^C during the first few iterations of the EM algorithm, depends on the size of the rating scale. For a large rating scale, the EM algorithm tends to remove quickly more components θ_{cl} during the first few iterations since the penalty term $\frac{N_r-1}{2}$ is high. On the other hand, for small

rating scales such as "accept" or "reject" patterns (i.e. $N_r = 2$), the model tends to maintain more classes (i.e. penalty $= 1/2$) to explain variable user ratings.

4 Experiments

We consider I-VCC-FMM and D-VCC-FMM as two variants of VCC-FMM where the visual features are represented in $\mathcal{V}$ and $\mathcal{X}$, respectively. By this way, we evaluate the impact on the prediction accuracy of the naive Bayes assumption among visual features. Two additional variants are also considered: V-FMM and V-GD-FMM. The former does not handle the contextual information and assumes θ^E_{ze} constant for all $e \in \mathcal{E}$. In the latter, feature selection is not considered by setting $\epsilon_{l_1} = 1$ and pruning uninformative components ξ_l for $l = 1, \ldots, D$.

4.1 Data Set

We have mounted an ASP.NET Web site with SQL Server database in order to collect ratings from 27 subjects who participated in the experiment (i.e. $N_u = 27$) during a period of two months. The participating subjects are graduate students in faculty of science. Subjects received periodically (twice a day) a list of three images on which they assign relevance degrees expressed on a five star rating scale (i.e. $N_r = 5$). We define the context as a combination of two attributes: location $\mathcal{L} = \{in-campus, out-campus\}$ and time as $\mathcal{T} = (weekday, weekend)$ i.e $N_e = 4$. After the period of rating's acquisition, a data set $\mathcal{D}$ of 11050 ratings is extracted from the SQL Server database (i.e. $N = 11050$). We have used a collection of 4775 (i.e. $N_v = 4775$) images collected in part from Washington University [1] and another part from collections of free photographs on the Internet. We have categorized manually this collection into 41 categories. 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. Then, 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") [6]. We have found that a number of 500 keypoints provided a good clustering for our collection. For global descriptors, we used the color correlogram [11] for image texture representation, and the edge histogram descriptor [12]. The color correlogram is built by considering the spatial arrangement of colors in the image for four displacements. A visual feature vector is represented in a 540-dimensional space ($D = 500 + 9 * 4 + 4 = 540$). We subdivide the data set $\mathcal{D}$ many times into two parts: for training and validation. Then, we measure the accuracy of the rating's prediction by the Mean Absolute Error (MAE) which is the average of the absolute deviation between the actual ratings (validation part) and the predicted ones.

[1] http://www.cs.washington.edu/research/imagedatabase.

4.2 First Experiment: Evaluating the Influence of Model Order on the Prediction Accuracy

This experiment investigates the influence of the assumed model order defined by K and M on the prediction accuracy of both I-VCC-FMM and D-VCC-FMM. While the number of image classes is known in the case of our collection, however, the number of user classes are not known in first sight. To validate the approach on a ground truth data $\mathcal{D}_{GT}$, we build a data set from preferences P_1 and P_2 of two most dissimilar subjects. We compute the dissimilarity in preferences on the basis of Pearson correlation coefficients. We sample ratings for 100 simulated users from the preferences P_1 and P_2 on images of four image classes. For each user, we generate 80 ratings ($\sim$ 20 ratings per context). Therefore, the ground truth model order is $K^* = 2$ and $M^* = 4$. The choice of image classes is purely motivated by convenience of presentation. Indeed, similar performance was noticed on the whole collection. We learn both I-VCC-FMM and D-VCC-FMM using one half of $\mathcal{D}_{GT}$ using different choices of training and validation data. The model order defined by $M = 15$ and $K = 15$ is used to initialize the EM algorithm for each partitioning of D_{GT}.

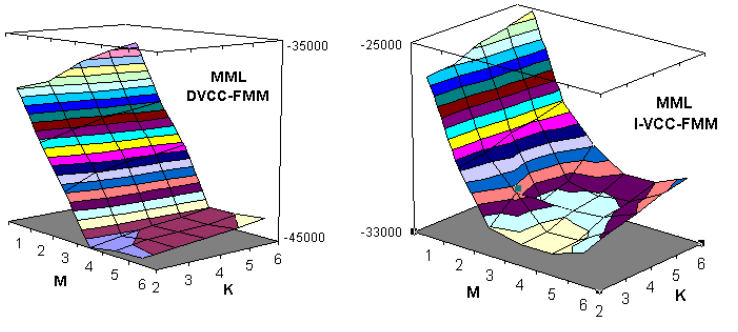


Fig. 3. MML criterion of the data set $\mathcal{D}_{GT}$ for D-VCC-FMM and I-VCC-FMM

Figure 3 shows that the MML approach has identified the correct number of user and image classes for both I-VCC-FMM and D-VCC-FMM on the synthetic data since the lowest MML was reported for the model order defined by $M = 4$ and $K = 2$. T The selection of the "correct" model order is important since it influences the accuracy of the prediction as illustrated by Figure 4. Furthermore, for $M > M^*$ the accuracy rating prediction is influenced more than the case of $K > K^*$. This experiment shows that the identification of the numbers of user and images classes is an important issue in CBIS.

4.3 Second Experiment: Comparison with State-of-the-Art

The aim of this experiment is to measure the contribution of the visual information and the user's context in making accurate predictions comparatively with some existing CF approaches. We make comparisons with the Aspect model [10],

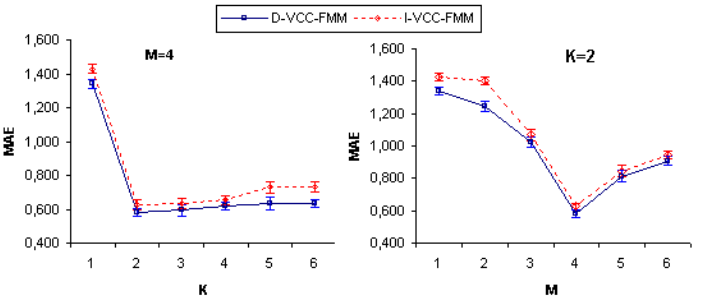
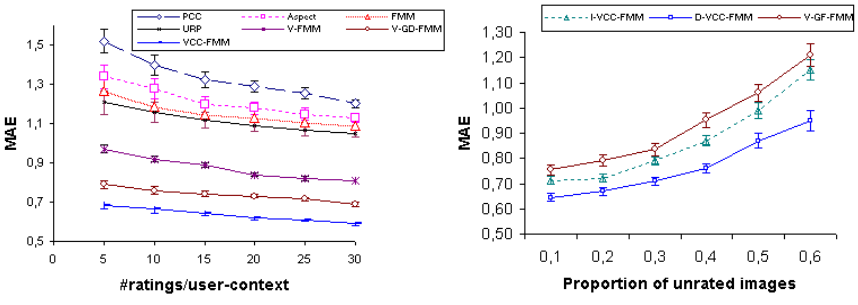


Fig. 4. Average MAE for different model orders

Pearson Correlation (PCC)[23], Flexible Mixture Model (FMM) [24], and User Rating Profile (URP) [17]. For accurate estimators, we learn the URP model using Gibs sampling. We retained for the previous algorithms, the model order that ensured the lowest MAE.

Table 1. Averaged MAE over 10 runs of the different algorithms on $\mathcal{D}$

	PCC(baseline)	Aspect	FMM	URP	V-FMM	V-GD-FMM	I-VCC	D-VCC
Avg MAE	1.327	1.201	1.145	1.116	0.890	0.754	0.712	0.645
Deviation	0.040	0.051	0.036	0.042	0.038	0.027	0.022	0.014
Improv.	0.00%	9.49%	13.71%	15.90%	32.94%	43.18%	51.62%	55.84%



(a) Data sparsity (b) new images

Fig. 5. MAE curves with error bars on the data set $\mathcal{D}$

The first five columns of table 1 show the added value provided by the visual information comparatively with pure CF techniques. For example, the improvement in the rating's prediction reported by V-FMM is 22.27% and 20.25% comparatively with FMM and URP, respectively. The algorithms (with context information) shown in the last three columns provide also an improvement (at least 15.28%) in the prediction accuracy comparatively with those which

do not consider the context of the user. Also, we notice that feature selection is another important factor due to the improvement provided by I-VCC-FMM (5.57%) and D-VCC-FMM (14.45%) comparatively with V-GD-FMM. Furthermore, the naive Bayes assumption in I-VCC-FMM has increased (10.39%) MAE of D-VCC-FMM. In addition, it is reported in figure 5(a) that VCC-FMM is less sensitive to data sparsity (number of ratings per user) than pure CF techniques. Finally, the evolution of the average MAE provided VCC-FMM for different proportions of unrated images remains under 25% for up to 30% of unrated images as shown in Figure 5(b). We explain the stability of the accuracy of VCC-FMM for data sparsity and new images by the visual information since only cluster representatives need to be rated.

5 Conclusions

In this paper, we have motivated the content based image suggestion by modeling long term user needs to the visual information. We have proposed a graphical model by addressing two issues of unsupervised learning: the feature selection and the model order identification. Experiments showed the importance of the visual information and the user's context in making accurate predictions.

References

1. Belk, R.W.: Situational Variables and Consumer Behavior. Journal of Consumer Research 2, 157–164 (1975)
2. Bouguila, N., Ziou, D.: A Hybrid SEM Algorithm for High-Dimensional Unsupervised Learning Using a Finite Generalized Dirichlet Mixture. IEEE Trans. on Image Processing 15(9), 1785–1803 (2006)
3. Bouguila, N., Ziou, D.: High-dimensional unsupervised selection and estimation of a finite generalized dirichlet mixture model based on minimum message length. IEEE Trans. on PAMI (2007)
4. Boutemedjet, S., Ziou, D.: Content-based collaborative filtering model for scalable visual document recommendation. In: Proc. of IJCAI-2007 Workshop on Multimodal Information Retrieval (2007)
5. Connor, R.J., Mosimann, J.E.: Concepts of Independence for Proportions With a Generalization of the Dirichlet Distribution. Journal of the American Statistical Association 39, 1–38 (1977)
6. Csurka, G., Dance, C.R., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3024, Springer, Heidelberg (2004)
7. Dy, J.G., Brodley, C.E.: Feature selection for unsupervised learning. Journal of Machine Learning Research 5, 845–889 (2004)
8. Figueiredo, M.A.T., Jain, A.K.: Unsupervised learning of finite mixture models. IEEE Trans. on PAMI 24(3), 4–37 (2002)
9. Guyon, I., Elisseeff, A.: An Introduction to Variable and Feature Selection. Journal of Machine Learning Research 3, 1157–1182 (2003)
10. Hofmann, T.: Probabilistic Latent Semantic Indexing. In: Proc. of SIGIR (1999)

11. Huang, J., Kumar, S.R., Mitra, M., Zhu, W.J., Zabih, R.: Image indexing using color correlograms. In: Proc. of IEEE Conf, IEEE Computer Society Press, Los Alamitos (1997)
12. Jain, A., Vailaya, A.: Image retrieval using color and shape. Pattern Recognition 29(8), 1233–1244 (1996)
13. Kontkanen, P., Myllymki, P., Silander, T., Tirri, H., Grnwald, P.: On predictive distributions and bayesian networks. Statistics and Computing 10(1), 39–54 (2000)
14. Law, M.H.C., Figueiredo, M.A.T., Jain, A.K.: Simultaneous feature selection and clustering using mixture models. IEEE Trans. on PAMI, 26(9) (2004)
15. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
16. Muramastsu, J., Pazzani, M., Billsus, D.: Syskill and Webert: Identifying Interesting Web Sites. In: Proc. of AAAI (1996)
17. Marlin, B.: Modeling User Rating Profiles For Collaborative Filtering. In: Proc. of NIPS (2003)
18. Messaris, P.: Visual Persuasion: The Role of Images in Advertising. Sage Pubns (1997)
19. Mooney, R.J., Roy, L.: Content-Based Book Recommending Using Learning for Text Categorization. In: Proc. 5th ACM Conf. Digital Libaries, ACM Press, New York (2000)
20. Ng, A.Y.: On feature selection: Learning with exponentially many irrelevant features as training examples. In: Proc. of ICML (1998)
21. Novovicova, J., Pudil, P., Kittler, J.: Divergence based feature selection for multimodal class densities. IEEE Trans. on PAMI 18(2), 218–223 (1996)
22. Popescul, A., Ungar, L.H., Pennock, D.M., Lawrence, S.: Probabilistic Models for Unified Collaborative and Content-Based Recommendation in Sparse-Data Environments. In: Proc. of UAI (2001)
23. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J.: Grouplens: An Open Architecture for Collaborative Filtering of Netnews. In: Proc. of ACM Conference on CSCW, ACM Press, New York (1994)
24. Si, L., Jin, R.: Flexible Mixture Model for Collaborative Filtering. In: Proc. of ICML, pp. 704–711 (2003)
25. Ueda, N., Nakano, R.: Deterministic Annealing EM Algorithm. Neural Networks 11(2), 271–282 (1998)
26. Vaithyanathan, S., Dom, B.: Generalized Model Selection for Unsupervised Learning in High Dimensions. In: Proc. of NIPS, pp. 970–976 (1999)
27. Wallace, C.: Statistical and Inductive Inference by Minimum Message Length. Information Science and Statistics. Springer, Heidelberg (2005)