Learning on Relevance Feedback in Content-based Image Retrieval

Hoi, Chu-Hong (Steven)

Supervisor:	Prof. Michael R. Lyu
Venue:	RM1027
Date:	11:00a.m. – 12:30p.m. 4 June, 2004

Outline

- 1. Introduction
- 2. Background & Related Work
- 3. Relevance Feedback with Biased SVM
- 4. Optimizing Learning with SVM Constraint
- 5. Group-based Relevance Feedback
- 6. Log-based Relevance Feedback
- 7. An Application: Web Image Learning
- 8. Discussions
- 9. Conclusions

1. Introduction

- Content-based Image Retrieval
- Relevance Feedback
- Contributions and Overview

Content-based Image Retrieval

- Visual information retrieval has been one of the most important and imperative tasks in computer science communities.
- CBIR is one of the most active and challenging research topics in visual information retrieval.
- Major research focuses in CBIR
  - Feature Identification and Representation
  - Distance Measure
  - Relevance Feedback Learning
  - Others (such as, database indexing issues, etc.)
- Challenges
  - A semantic gap between low-level features and high-level concepts
  - Subjectivity of human perception of visual content
  - Others (such as semantic understanding, annotation, clustering, etc…)

Relevance Feedback in CBIR

- Relevance feedback is a powerful technique to bridge the semantic gap of CBIR and overcome the subjectivity of human perception of visual content.
- Although many techniques has been proposed, existing methods have many drawbacks and limitations, particularly in the following aspects:
  - most without noticing the imbalanced dataset problem
  - paying less attention on the insufficient training samples
  - normally assuming samples are drawn from one positive class and one negative class
  - typically requiring a lot of rounds of feedback in order to achieve satisfactory results

Contributions and Overview

- Relevance Feedback with Biased SVM
  - Addressing the imbalance problem of relevance feedback
  - Proposing Biased SVM to construct the relevance feedback algorithm for attacking the imbalance problem
- Optimizing Learning (OPL) with SVM Constraint
  - Attacking insufficient training samples
  - Unifying OPL and SVM for learning similarity measure
- Group-based Relevance Feedback Using SVM Ensembles
  - Relaxing the assumption of regular relevance feedback: the training samples of relevance feedback are based on (x+1)-class model
  - Constructing a novel and effective group-based relevance feedback algorithm using SVM ensembles

Contributions and Overview (cont.)

- Log-based Relevance Feedback Using Soft Label SVM
  - Studying the techniques for learning user feedback logs
  - Proposing a modified SVM for log-based relevance feedback algorithms
- Application: Web Image Learning for Searching Semantic Concepts in Image Databases
  - Suggesting a novel application for learning semantic concepts by Web images in image databases
  - Employing a relevance feedback mechanism to attack the learning tasks
- *Other related work on multimedia retrieval*
  - Video similarity detection, face recognition using MPM

Outline

- 1. Introduction
- 2. Background & Related Work
- 3. Relevance Feedback with Biased SVM
- 4. Optimizing Learning with SVM Constraint
- 5. Group-based Relevance Feedback
- 6. Log-based Relevance Feedback
- 7. An Application: Web Image Learning
- 8. Discussions
- 8. Conclusions

2. Background & Related Work

- Relevance Feedback in CBIR
  - Problem Statement
  - Related Work
    - Heuristic weighting scheme
    - Optimal Formulations
    - Varied Machine Learning Techniques
- Support Vector Machines
  - Basic learning concepts
  - The optimal separating hyperplane
  - nu-SVM and 1-SVM

Relevance Feedback in CBIR

- Problem Statement
  - Definition
    Relevance feedback is the process of automatically altering an existing query employing information provided by users about the relevance of previous retrieved objects in order to approach the users' query targets.
  - Steps
    - Step 1: Init query: Query-by-Example (or by keywords, random seeds)
    - Step 2: Judge relevance on the retrieved results: relevant/irrelevant Relevant samples are regarded as "positive" data, while irrelevant ones are "negative".
    - Step 3: Learn with the fed-back information and return the results
    - Step 4: Repeat step 2 until the users find their targets

Relevance Feedback in CBIR (cont.)

- Related Work
  - Heuristic Weighting Schemes
    - Query modification: query point movement, query expansion
    - Query re-weighting
  - Optimization Formulations
    - Formulating the task as an optimization problem: Mindreader
    - More rigorous and systematical based on hierarchical models: Optimizing Learning (OPL)
  - Varied Machine Learning Techniques
    - Support Vector Machine (SVM)
    - Others: Neural Networks, Decision Tree, etc.

Support Vector Machines

- Basic Learning Concepts
  - We consider the learning problem as a problem of finding a desired dependence using a limited number of observations.
  - Two inductive learning principles
    - Empirical Risk Minimization (ERM): minimizing error on training data
    - Structural Risk Minimization (SRM): minimizing bounds of risk on test data
  - SVM is a large margin learning algorithm that implements the SRM principle.

Support Vector Machines

- The optimal Separating hyperplane

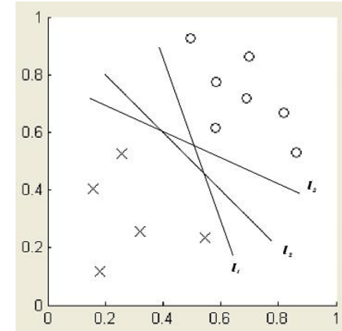
$$\min_{\mathbf{w},b} \quad \frac{1}{2}\|\mathbf{w}\|^2$$
$$s.t. \quad y_i((\mathbf{w}\cdot\mathbf{x}_i)+b)\geq 1, \quad i=1,\ldots,l$$

- nu-SVM (soft-margin & kernel)

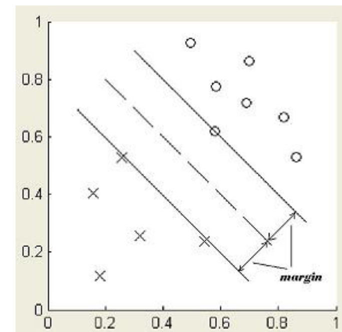
$$\min_{\mathbf{w}} \quad \frac{1}{2}\|\mathbf{w}\|^2-\nu\rho+\frac{1}{l}\sum\xi_i \tag{2.14}$$
$$s.t. \quad y_i(\mathbf{w}\cdot\Phi(\mathbf{x}_i)+b)\geq\rho-\xi_i \tag{2.15}$$
$$\xi_i\geq 0, \rho\geq 0, \quad i=1,\ldots,l, \tag{2.16}$$

- One-class SVM (1-SVM)

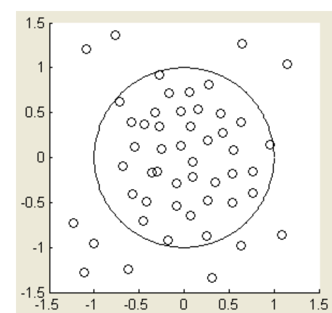
$$\min_{R,\mathbf{c}} \quad R^2+\frac{1}{\nu l}\sum_i\xi_i \tag{2.17}$$
$$s.t. \quad \|\Phi(\mathbf{x}_i)-\mathbf{c}\|^2\leq R^2+\xi_i, \tag{2.18}$$
$$\xi_i\geq 0 \quad i=1,\ldots,l\,. \tag{2.19}$$



(a)



(b)



(c)

Outline

- 1. Introduction
- 2. Background & Related Work
- 3. Relevance Feedback with Biased SVM
- 4. Optimizing Learning with SVM Constraint
- 5. Group-based Relevance Feedback
- 6. Log-based Relevance Feedback
- 7. An Application: Web Image Learning
- 8. Conclusions
- 9. Q&A

3. Relevance Feedback with Biased SVM

- Motivation
  - The imbalance problem of relevance feedback
    - Negative samples normally outnumber positive samples.
    - Positive samples are clustered in the same way while negative samples are positioned in the different ways.
    - Problem/risk: the positive samples are easily overwhelmed by the negative samples in regular learning algorithm without bias consideration.
  - Related Work
    - Regular two-class SVM-based relevance feedback simply regards the problem as a pure two-class classification task.
    - Relevance feedback with regular 1-SVM seems avoid the problem. However, the relevance feedback job can be done well if the negative information is ignored.
  - Biased SVM, a modified 1-SVM technique, is proposed to construct the relevance feedback algorithm for attacking the imbalance problem of relevance feedback.

Biased SVM

- Problem Formulation
  - Let us consider the following training data:

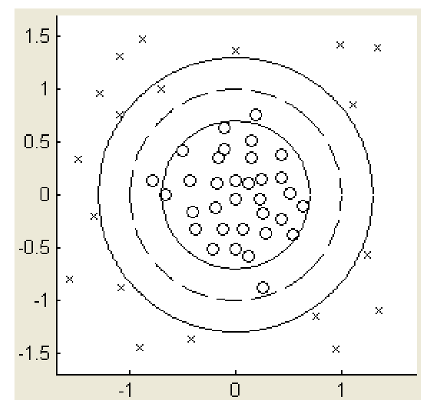
$$(x_1, y_1), \ldots, (x_l, y_l) \in \mathbb{R}^m \times Y, \quad Y = \{-1, +1\}$$

  - The goal of Biased SVM is to find the optimal hyper-sphere to classify the positive and negative samples.
  - The objective function

$$\min_{R\in\mathbb{R},\rho\in\mathbb{R},\mathbf{c}\in\mathcal{F}} \quad bR^2 - \rho + \frac{1}{\nu l}\sum_{i=1}^{l}\xi_i \qquad (3.2)$$

$$s.t. \quad y_i(||\Phi(\mathbf{x}_i) - \mathbf{c}||^2 - R^2) \leq -\rho + \xi_i\,, \qquad (3.3)$$

$$b > 0\,, \rho \geq 0\,, 0 \leq \nu \leq 1, \xi_i \geq 0\,, \qquad (3.4)$$



Biased SVM (cont.)

- Solution to the optimization problem
  - Introducing the Lagrangian:

$$L(R, \xi, c, \alpha, \beta, \lambda) = bR^2 - \rho + \frac{1}{\nu l}\sum_{i=1}^{l}\xi_i - \sum_{i=1}^{l}\beta_i\xi_i - \lambda\rho + \sum_{i=1}^{l}\alpha_i[y_i(||\Phi(\mathbf{x}_i) - \mathbf{c}||^2 - R^2) + \rho - \xi_i]\,.$$

  - Let us take the partial derivatives with L:

$$2R(b - \sum_{i=1}^{l} y_i\alpha_i) = 0 \;\Rightarrow\; \sum_{i=1}^{l} y_i\alpha_i = b\,; \qquad (3.6)$$

$$\frac{1}{\nu l} - \alpha_i - \beta_i = 0, \;\Rightarrow\; 0 \le \alpha_i \le \frac{1}{\nu l}\,; \qquad (3.7)$$

$$\sum_{i=1}^{l} 2\alpha_i y_i(\Phi(\mathbf{x}_i) - \mathbf{c}) = 0 \;\Rightarrow\; \mathbf{c} = \frac{1}{b}\sum_{i=1}^{l}\alpha_i y_i\Phi(\mathbf{x}_i) \qquad (3.8)$$

$$-1 + \sum_{i=1}^{l}\alpha_i - \lambda = 0 \;\Rightarrow\; \sum_{i=1}^{l}\alpha_i \ge 1\,. \qquad (3.9)$$

Biased SVM (cont.)

- The dual problem can be derived as:

$$\max_{\alpha} \quad \sum_i \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}_i) - \frac{1}{b} \sum_{i,j} \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j) \tag{3.10}$$

$$s.t. \quad \sum_i \alpha_i y_i = b \ , \tag{3.11}$$

$$0 \le \alpha_i \le \frac{1}{\nu l} \ , \tag{3.12}$$

$$\sum_i \alpha_i \ge 1, \quad i = 1, 2, \ldots, l \ . \tag{3.13}$$

- This can be solved by the Quadratic Programming technique.
- After solving the problem, we can obtain its decision function:

$$\begin{aligned} f(\mathbf{x}) &= R^2 - \|\Phi(\mathbf{x}) - \mathbf{c}\|^2 \\ &= R^2 - \|\Phi(\mathbf{x}) - \frac{1}{b} \sum_{i=1}^{l} \alpha_i y_i \Phi(\mathbf{x}_i)\|^2 \ , \end{aligned} \tag{3.18}$$

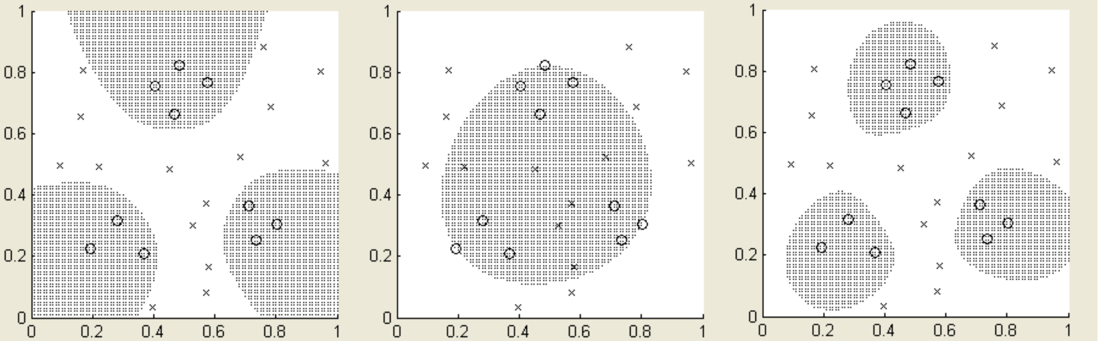
Relevance Feedback by BSVM

- Difference between Biased SVM, nu-SVM:

$$\sum_i \alpha_i y_i = b \qquad \sum_{i \in S^+} \alpha_i - \sum_{i \in S^-} \alpha_i = b$$

$$\sum_{i \in S^+} \alpha_i - \sum_{i \in S^-} \alpha_i = 0$$

- Visual comparison of three different approaches



(a) SVM (ν-SVM) (b) 1-SVM (c) BSVM

Figure 3.2: Decision boundaries of three classification methods with the same kernel (RBF) and parameters (γ=0.1).

Relevance Feedback by BSVM (cont.)

- The final decision function for BSVM

$$
\begin{aligned}
f(\mathbf{x}) &= R^2 - ||\Phi(\mathbf{x}) - \mathbf{c}||^2 \\
&= R^2 - ||\Phi(\mathbf{x}) - \frac{1}{b}\sum_{i=1}^{l} \alpha_i y_i \Phi(\mathbf{x}_i)||^2 \, , \qquad (3.18)
\end{aligned}
$$

- For relevance feedback tasks, we can simply employ the following function to rank the samples

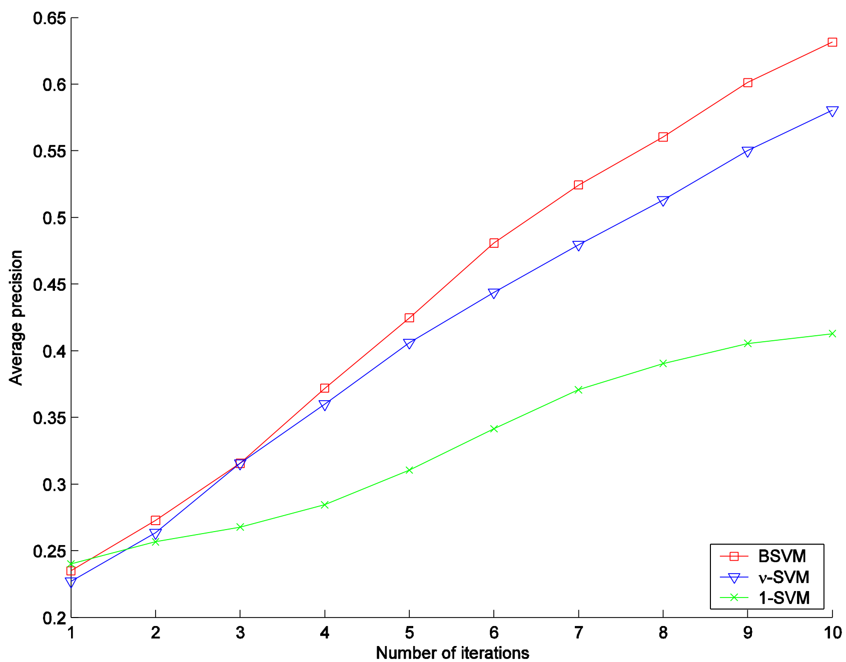
$$
f(\mathbf{x}) = \frac{2}{b}\sum_{i} \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) - k(\mathbf{x}, \mathbf{x}) \, . \qquad (3.19)
$$

Experiments

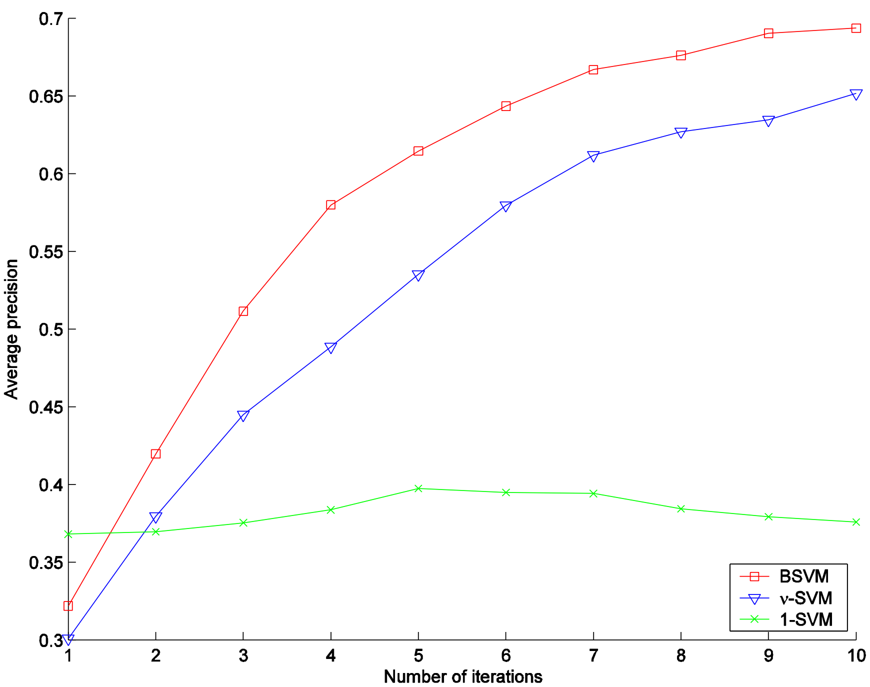
- Datasets
  - One synthetic dataset: 40-Cat, each contains 100 data points randomly generated by 7 Gaussian in a 40-dimensional space. Means and covariance matrices of the Gaussians in each category are randomly generated in the range of [0,10].
  - Two real-world image datasets selected from COREL image CDs
    - 20-Cat: 2,000 images
    - 50-Cat: 5,000 images
- Image Representation
  - Color Moment (9-dimension)
  - Edge Direction Histogram (18-dimension, Canny detector, 18 bins of 20 degrees)
  - Wavelet-based texture (9-dimension, Daubechies-4 wavelet, 3-level DWT, 9 subimages are selected to generate the feature)
- Compared Schemes
  - Relevance Feedback with nu-SVM
  - Relevance Feedback with 1-SVM
  - Relevance Feedback with BSVM

Experiments (cont.)

- Experimental results



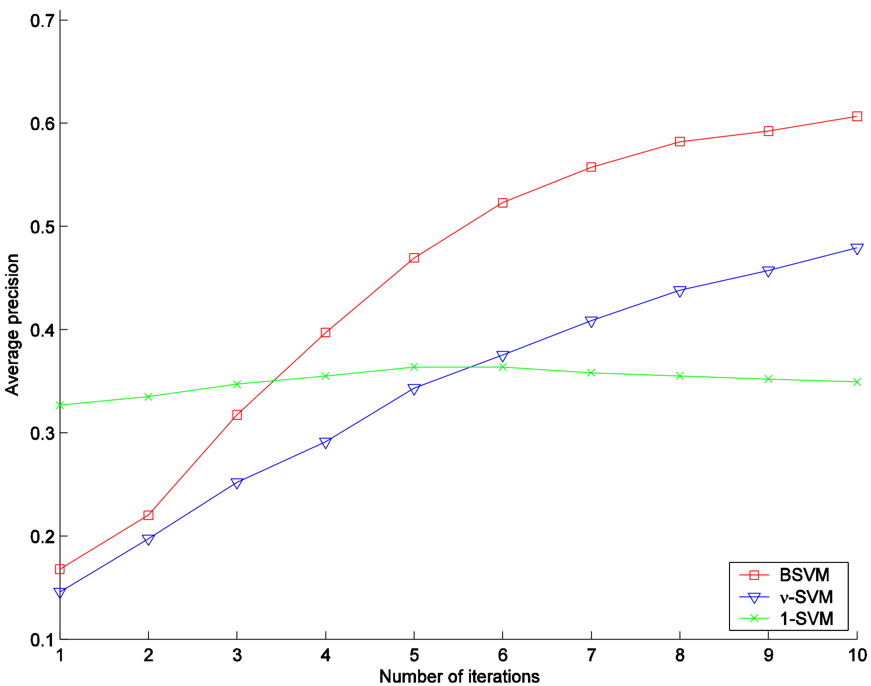
Synthetic dataset



20-Cat COREL Images

Experiments (cont.)

- Experimental results



50-Cat COREL Images

Table 3.1: Average precision after 10 iterations

Methods	Top20@20-Cat	Top30@20-Cat	Top50@20-Cat
ν-SVM	0.656	0.648	0.608
1-SVM	0.401	0.396	0.346
BSVM	0.713	0.694	0.650
Methods	Top20@50-Cat	Top30@50-Cat	Top50@50-Cat
ν-SVM	0.487	0.480	0.465
1-SVM	0.376	0.358	0.344
BSVM	0.639	0.614	0.588

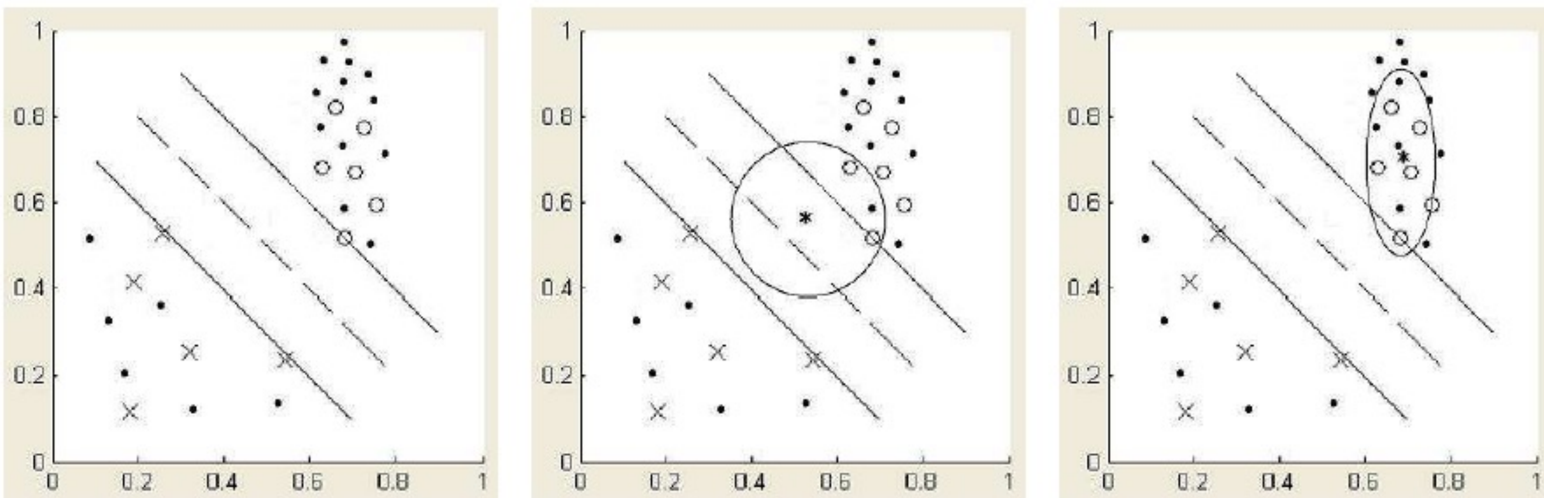
Outline

- 1. Introduction
- 2. Background & Related Work
- 3. Relevance Feedback with Biased SVM
- 4. Optimizing Learning with SVM Constraint
- 5. Group-based Relevance Feedback
- 6. Log-based Relevance Feedback
- 7. An Application: Web Image Learning
- 8. Discussions
- 9. Conclusions

4. Optimizing Learning with SVM Constraint

- Motivation
  - Learning optimal distance measure by relevance feedback is a challenging problem in CBIR.
  - Two important relevance feedback techniques
    - Optimizing Learning (OPL)
    - SVM-based Learning
  - Limitation of OPL
    - It does not support kernel-based learning.
    - Its performance is not competitive with kernel techniques.
  - Limitation of SVM
    - Inaccurate boundary when facing insufficient training samples
    - Ranking the samples simply employing the distance from boundary may not be effective when facing the inaccurate boundary.
  - Key idea
    - Unify the OPL and SVM techniques, first employing SVM to classify the samples, and then combining OPL to learn and rank the samples based on the boundary of SVM
    - The optimal distance measure learned with the OPL by the SVM constraint will be more effective and sophisticated when facing insufficient training samples.

- **Motivation (cont.)**
  - Comparison of different approaches



(a) SVM (b) SVM+EU (c) SVM+OPL

- Problem formulation
  - Goal: learning an optimal distance function
  - Notations (details)
    - $\mathcal{D}_{SVM}()$ SVM distance – Coarse distance
    - $\mathcal{D}_{OPL}()$ OPL distance – Fine distance
    - $\mathbf{Dis}()$ Overall distance measure unifying SVM & OPL
  - Procedures of the learning scheme
    - 1. Learn the classification boundary by SVM
    - 2. Learn the distance function by OPL with the SVM constraint
    - 3. The overall distance function is unified with OPL and SVM. The samples inside the boundary of SVM are ranked by the OPL distance, otherwise they are ranked by the SVM distance.

- Learning the boundary by SVM

$$\min_{\mathbf{w}\in\mathcal{F}} \quad \frac{1}{2}\|\mathbf{w}\|^2 - \nu\rho + \frac{1}{N}\sum \xi_i \qquad (4.2)$$
$$s.t. \quad y_i(\mathbf{w}\cdot\Phi(\mathbf{x}_i)) \geq \rho - \xi_i \qquad (4.3)$$
$$\xi_i \geq 0, \rho \geq 0, \quad i = 1, \ldots, N, \qquad (4.4)$$

$$\mathcal{D}_{SVM}(\mathbf{x_n}, \Theta) = \sum_{\mathbf{x_i}\in S} \alpha_i y_i K(\mathbf{x_i}, \mathbf{x_n}) + b$$

- Optimal distance measure by OPL
  - Straight Euclidean Distance $d_n = (\mathbf{q} - \mathbf{x_n})^T(\mathbf{q} - \mathbf{x_n})$
  - Generalized Ellipsoid Distance (GED) $d_n = (\mathbf{q} - \mathbf{x}_n)^T W(\mathbf{q} - \mathbf{x_n})$
    where ***W*** is a real symmetric full matrix
  - The distance measure by GED: $D(\mathbf{x_n}, \mathbf{q}) = \sum_{i=1}^{M} u_i(\mathbf{x_{ni}} - \mathbf{q_i})^T W_i(\mathbf{x_{ni}} - \mathbf{q_i})$
  - The parameters to be optimized: **q**, ***W***, **u**

Optimal distance measure by OPL (cont.)

- The objective of optimization

$$\min_{W,\mathbf{q},\mathbf{u}} \quad \sum_{n=1}^{N}\sum_{i=1}^{M} v_n u_i (\mathbf{x_{ni}} - \mathbf{q_i})^T W_i (\mathbf{x_{ni}} - \mathbf{q_i}) \tag{4.14}$$

$$\text{s.t.} \quad \sum_{i=1}^{M} \frac{1}{u_i} = 1 \tag{4.15}$$

$$det(W_i) = 1 \quad i = 1, 2, \ldots, M\ . \tag{4.16}$$

$$v(\mathbf{x_i}) = \frac{exp(D_{SVM}(\mathbf{x_i}, \Theta))}{1 + exp(D_{SVM}(\mathbf{x_i}, \Theta))}$$

- The solutions to the problem

$$\mathbf{q_i}^* = \frac{X_i^T \mathbf{v}}{\sum_{n=1}^{N} v_n} \qquad W_i^* = \begin{cases} (det(C_i))^{\frac{1}{L_i}} C_i^{-1}, & N \geq L_i \\ diag(\frac{1}{\sigma_1^2}, \ldots, \frac{1}{\sigma_{L_i}^2}), & N < L_i \end{cases}$$

$$u_i^* = \sum_{j=1}^{M} \sqrt{\frac{f_j}{f_i}} \qquad C_{ist} = \frac{\sum_{n=1}^{N} v_n (\vec{x}_{nis} - \vec{q}_{is})(\vec{x}_{1it} - \vec{q}_{it})}{\sum_{n=1}^{N} v_n}$$

$$f_i = \sum_{n=1}^{N} v_n (\vec{x_{ni}} - \vec{q_i})^T W_i (\vec{x_{ni}} - \vec{q_i})$$

- Overall Dissimilarity Measure Unifying OPL and SVM

$$\mathbf{Dis}(\mathbf{x_n}) = \begin{cases} \mathcal{D}_{OPL}(\mathbf{x_n}, \mathbf{q}^*), & D_{SVM}(\mathbf{x_n}, \Theta) \geq 0 \\ MaxDis - \mathcal{D}_{SVM}(\mathbf{x_n}, \Theta), & \mathcal{D}_{SVM}(\mathbf{x_n}, \Theta) < 0 \end{cases} \quad (4.21)$$

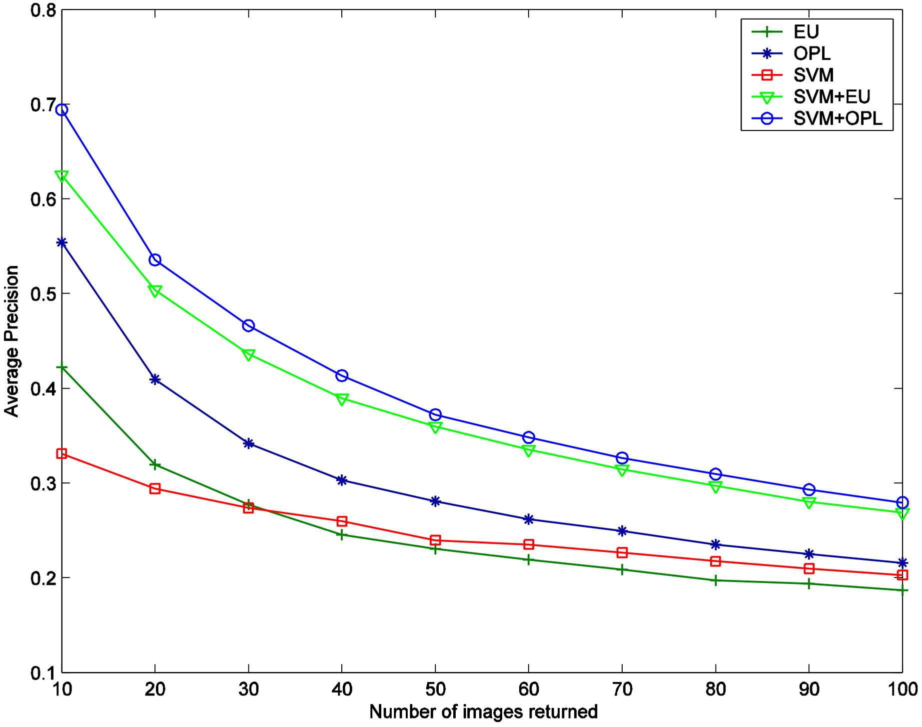
$$MaxDis = \max_{\forall x_i} D_{OPL}(\mathbf{x_i}, \mathbf{q}^*), \quad \text{if } D_{SVM}(\mathbf{x_i}, \Theta) \geq 0 \; . \qquad (4.22)$$

Experiments

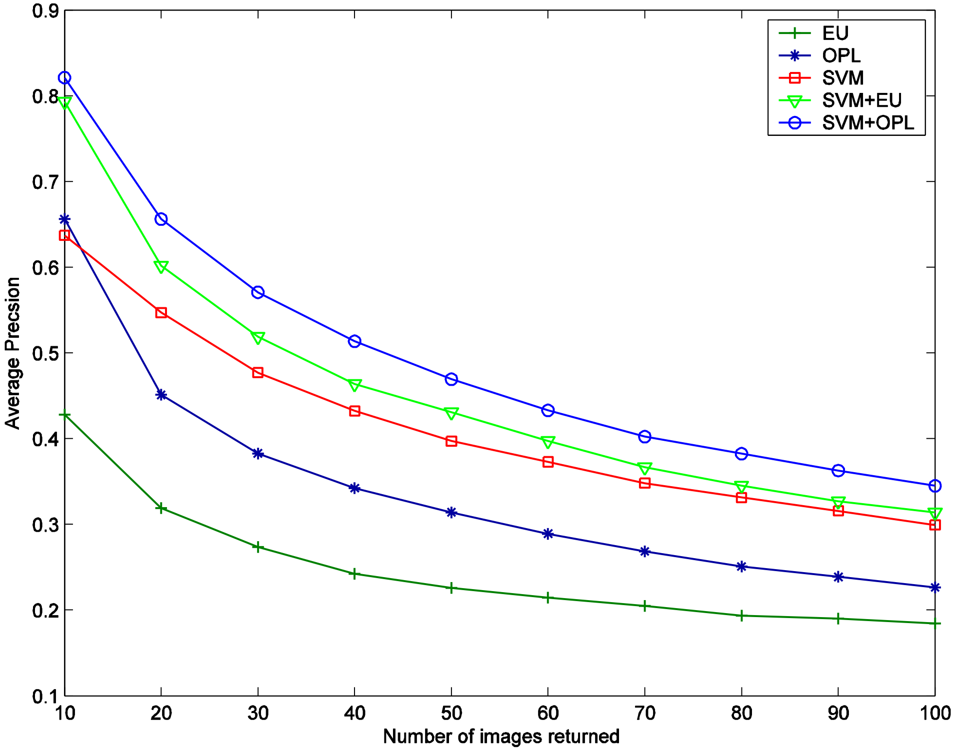
- Datasets
  - Natural images are selected from COREL CDs to form two datasets:
    - 20-Category: 2,000 images
    - 50-Category: 5,000 images
- Image Representation
  - Color Moment (9-dimension)
  - Edge Direction Histogram (18-dimension)
  - Wavelet-based Texture (9-dimension)
- Experimental Parameters
  - Radial Basis Function (RBF) Kernel for SVMs
- Schemes for comparison
  - EU (Euclidean distance)
  - OPL (Optimizing Learning)
  - SVM
  - SVM+EU
  - SVM+OPL

Experiments (cont.)

- Experimental results on the 20-Cat dataset



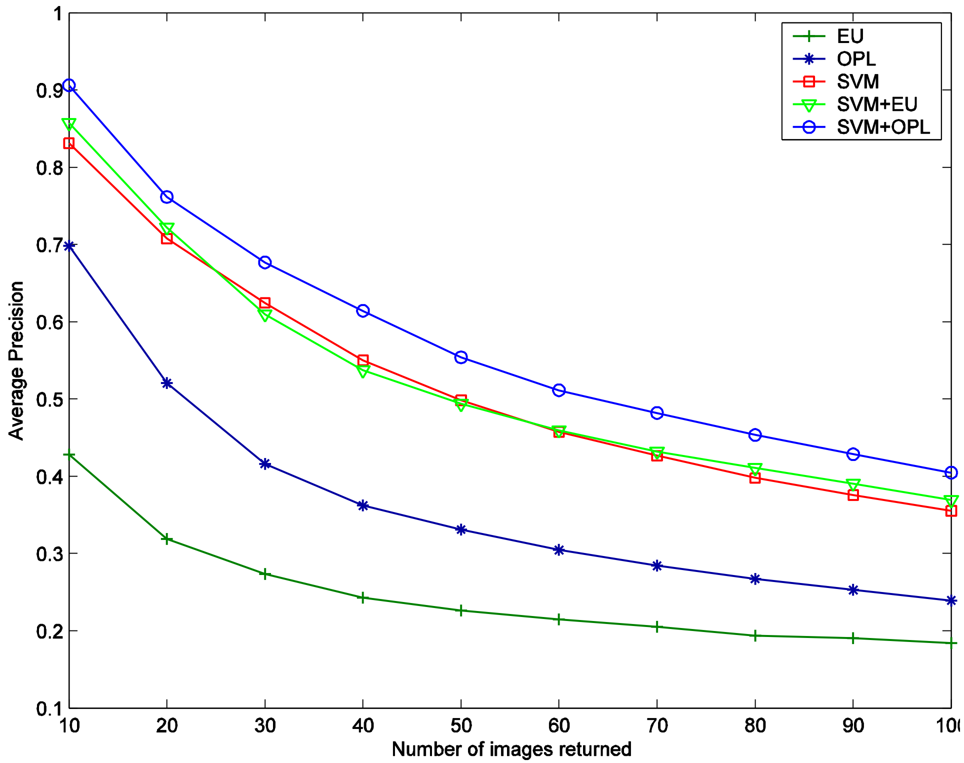
Round 1



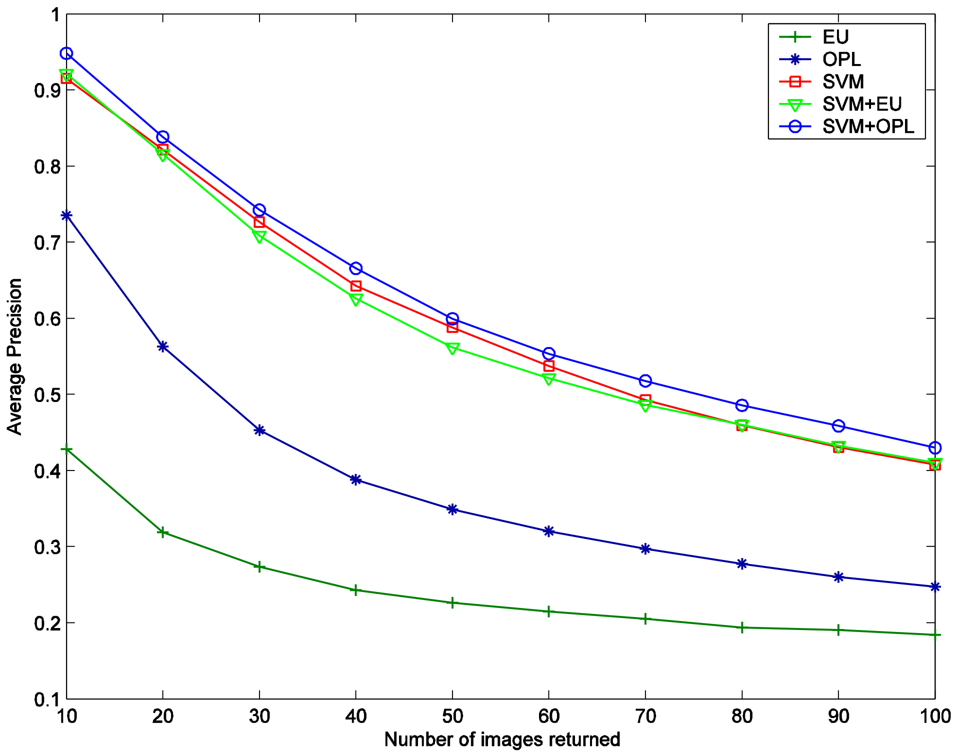
Round 2

Experiments (cont.)

- Experimental results on the 20-Cat dataset



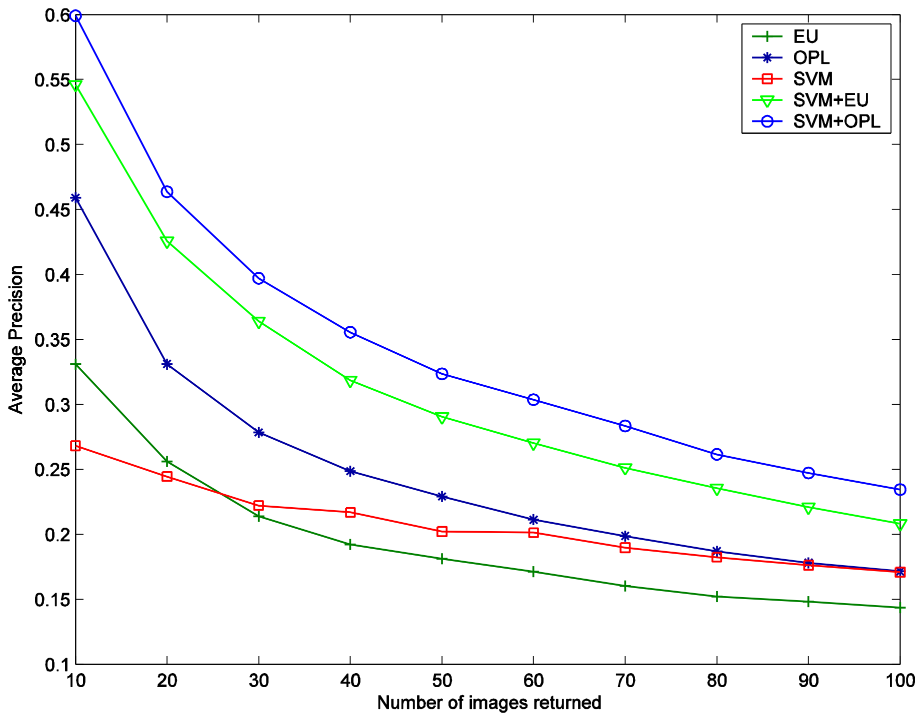
Round 3



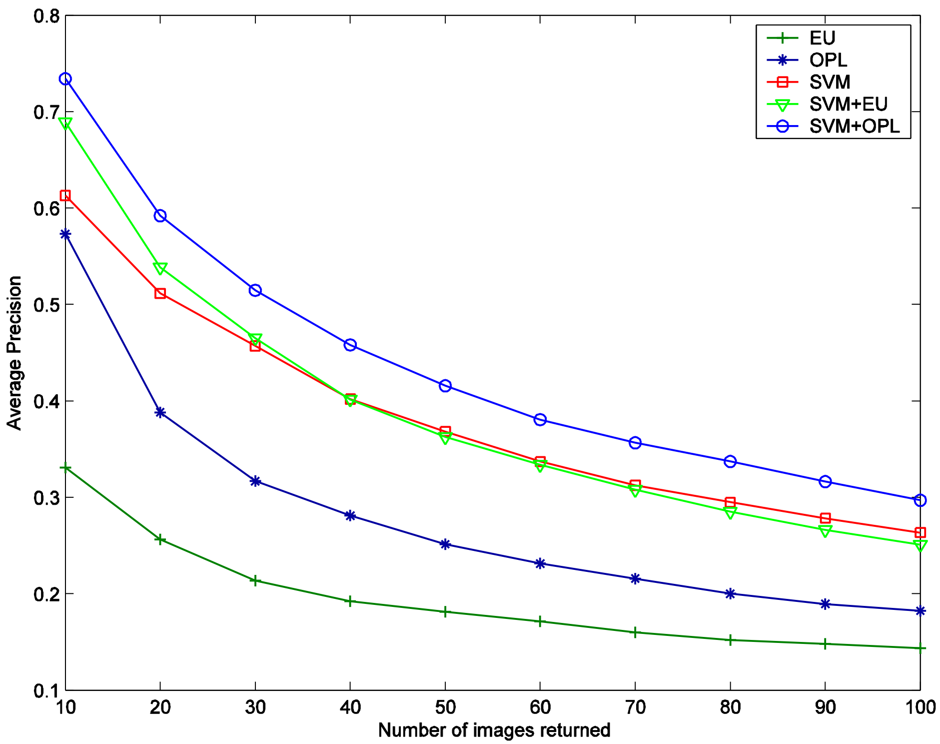
Round 4

Experiments (cont.)

- Experimental results on the 50-Cat dataset



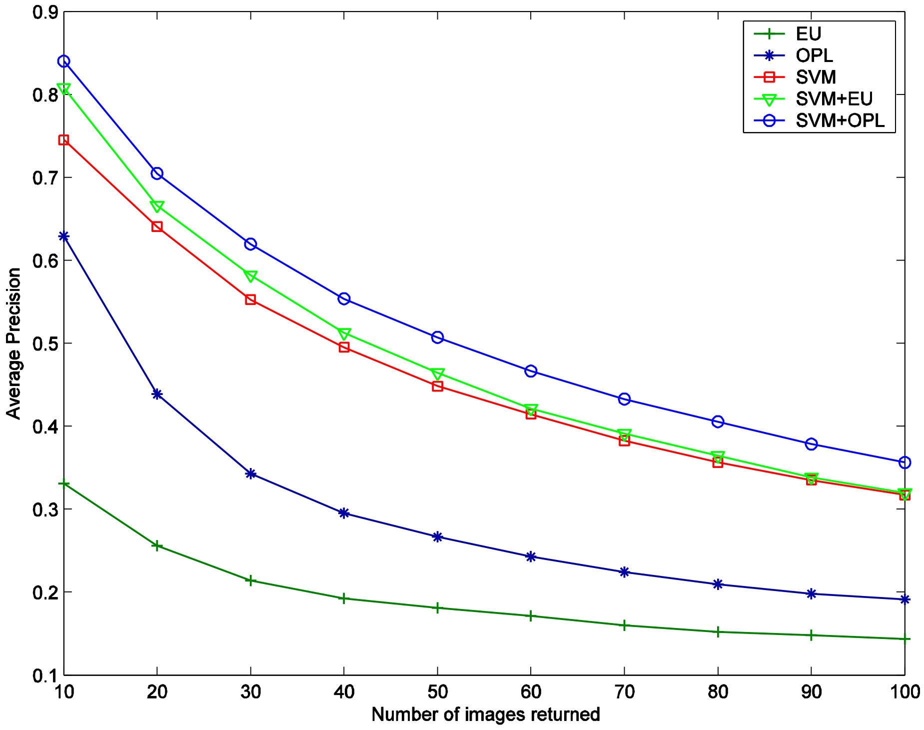
Round 1



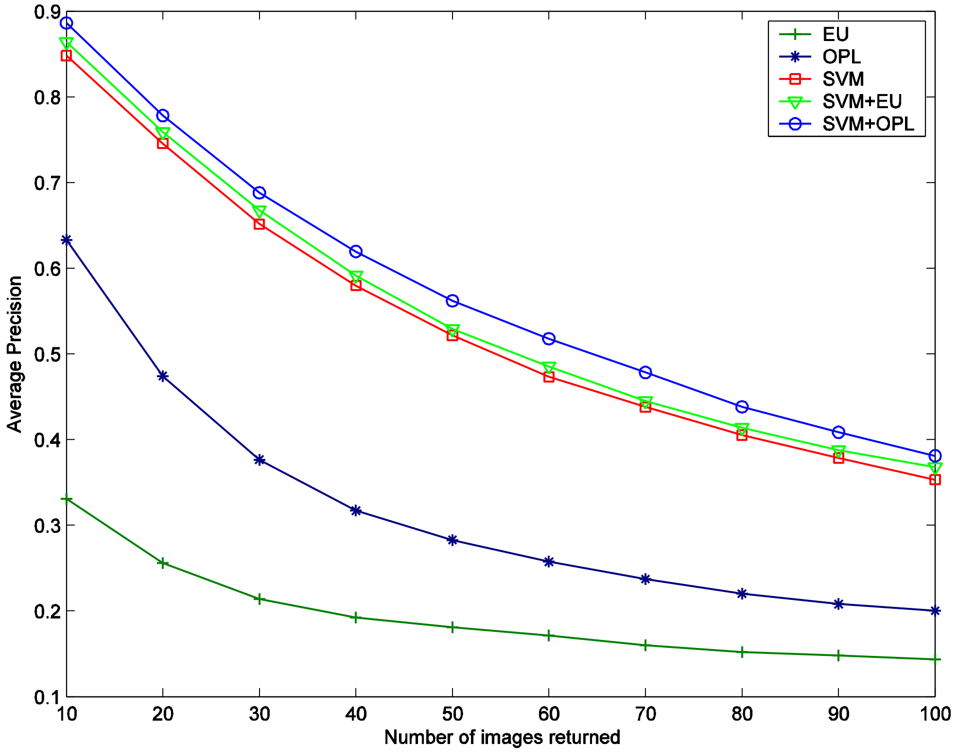
Round 2

Experiments (cont.)

- Experimental results on the 50-Cat dataset



Round 3



Round 4

Experiments (cont.)

- Time Complexity Performance

Table 4.2: Time cost of our proposed approach

Dataset	Size	$T_{solving\ OPL}$	$T_{SVM\ Training}$	$T_{all\ executions}$
20-Cat	2000	2.78 ± 0.05	5.81 ± 0.09	49.84 ± 0.11
50-Cat	5000	2.90 ± 0.05	6.39 ± 0.07	68.75 ± 0.08

For 100 executions
in average, less than 0.2 second for one feedback round

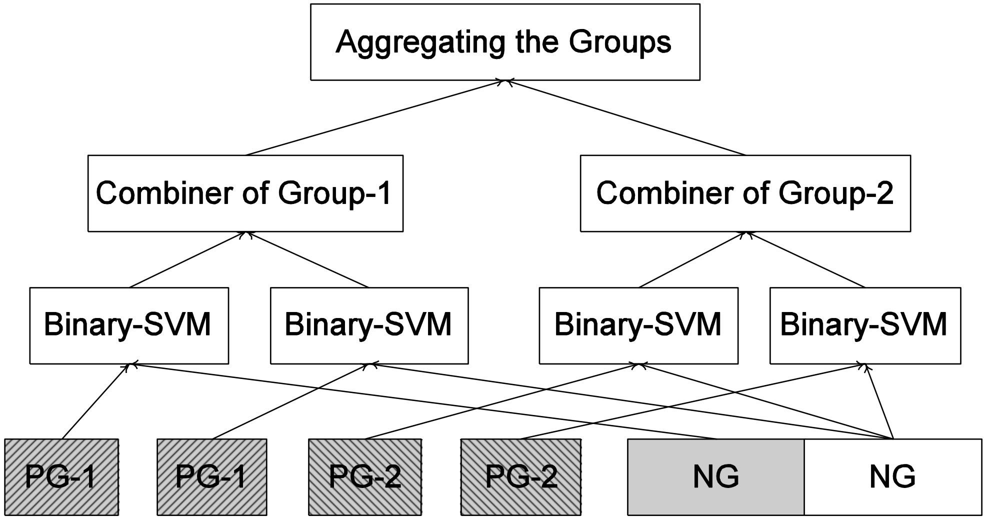
Outline

- 1. Introduction
- 2. Background & Related Work
- 3. Relevance Feedback with Biased SVM
- 4. Optimizing Learning with SVM Constraint
- 5. Group-based Relevance Feedback
- 6. Log-based Relevance Feedback
- 7. An Application: Web Image Learning
- 8. Discussions
- 9. Conclusions

5. Group-based Relevance Feedback

- Motivation
  - Class assumption: regular approaches typically regard the data of relevance feedback are drawn from one positive class and one negative class.
  - Problem: not effective enough to describe the data
  - Other related Work:
    - (1+x)-class assumption
      - One positive class and multiple negative classes
    - (x+y)-class assumption
      - Multiple positive classes and multiple negative classes
  - Our (x+1)-class assumption
    - Multiple positive classes and one negative class
    - Users are more interested in relevant samples rather than the irrelevant ones.
    - More practical and effective than regular approaches
  - We suggest to "group" the positive samples and propose a group-based relevance feedback algorithm using SVM ensembles

- Proposed Architecture
  - SVM Ensembles
    - A collection of several SVM classifiers
  - Constructing Method
    - Group the positive samples by users
    - The negative samples are partitioned to several parts which are formed with the positive group for training each SVM classifier
  - A figure illustrates an example of the proposed architecture



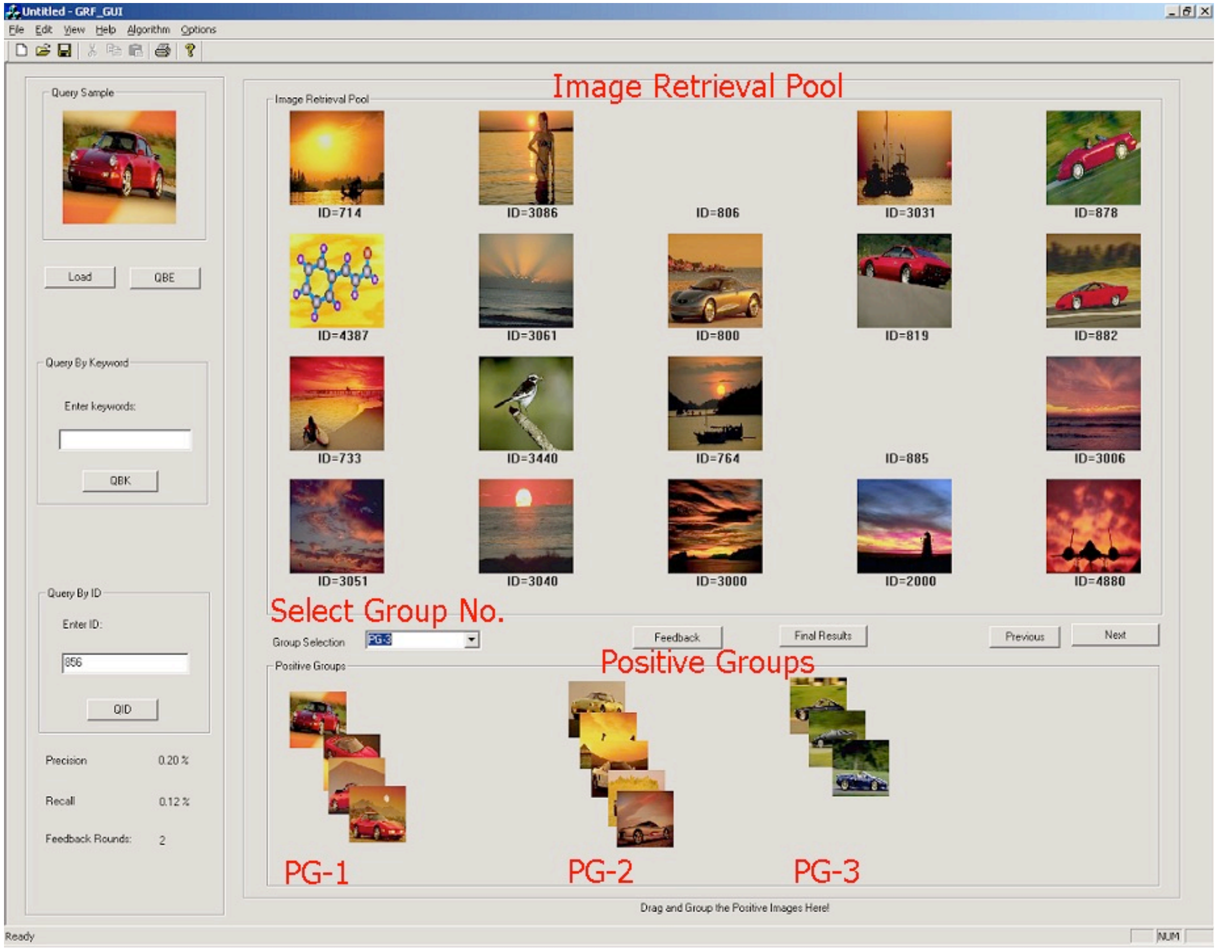
- Proposed Architecture
  - Notations
    - K_g – number of positive groups
    - K_m – number of SVM classifiers in each positive group
    - f_{ij} – the decision function of the j-th SVM in the i-th ensemble
  - Strategy for combination and Group Aggregation
    - Based on Sum Rule and linear combination with weights
  - The final decision function $f_{GRF}(\mathbf{x})$ is given as

$$F_i(\mathbf{x}) = \sum_{j=1}^{K_m} f_{ij}(\mathbf{x})$$

$$f_{GRF}(\mathbf{x}) = \sum_{i=1}^{K_g} w_i F_i(\mathbf{x}) = \sum_{i=1}^{K_g} \sum_{j=1}^{K_m} w_i f_{ij}(\mathbf{x})$$

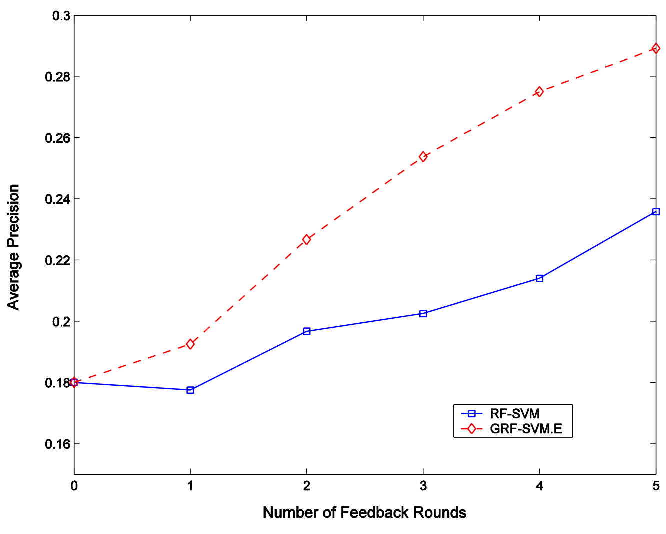
Experiments

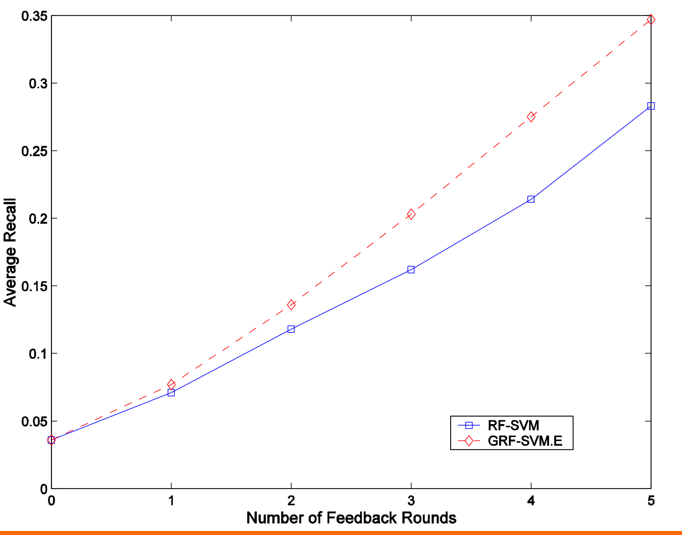
- The CBIR System for Group Evaluation



Experiments (cont.)

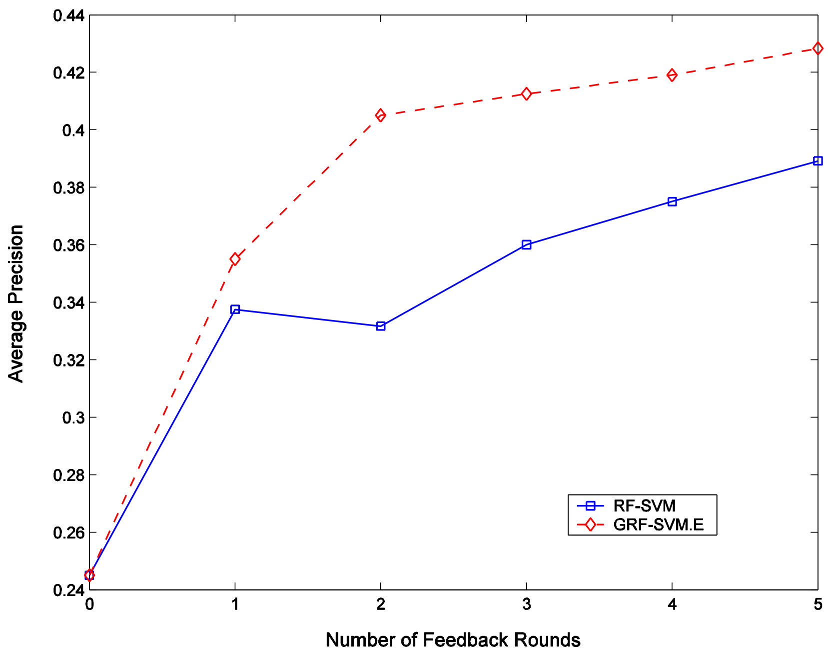
- Experimental results
  - Test database: 50 Categories of images
  - Features: color moment, edge direction histogram, DWT texture
  - Kernel: RBF
  - 5 rounds of feedback, 20 images each round
- Retrieval Performance for searching “cars”

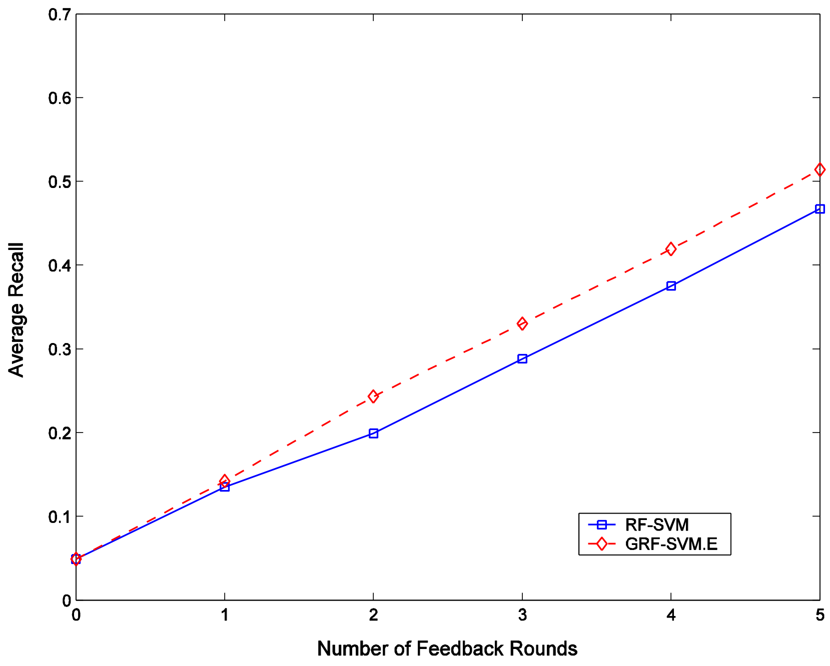




Experiments (cont.)

- Retrieval Performance for searching "roses"





Outline

- 1. Introduction
- 2. Background & Related Work
- 3. Relevance Feedback with Biased SVM
- 4. Optimizing Learning with SVM Constraint
- 5. Group-based Relevance Feedback
- 6. Log-based Relevance Feedback
- 7. An Application: Web Image Learning
- 8. Discussions
- 9. Conclusions

6. Log-based Relevance Feedback

- Motivation
  - In regular relevance feedback, retrieval results of the initial rounds of feedback are not very good.
  - Users typically are required to do a lot of rounds of feedback in order to achieve satisfactory results.
  - In a long-term study purpose, we suggest to employ the user feedback logs to improve the regular relevance feedback tasks.
  - To engage users' logs, we proposed a modified SVM technique called Soft Label SVM to formulate the relevance feedback algorithm.

- Problem formulation
  - A Relevance Matrix (RM) is constructed by the feedback logs to represent the relevance relationship between images.
  - Suppose image *i* is marked as relevant and *j* is marked as irrelevant in a given session *k*, then
    RM (*k*, *i*) = 1 and RM (*k*, *j*) = -1
  - The relationship of two images *i* and *j* can be expressed as

$$R_{ij} = RM(:,i)^T \cdot RM(:,j) \ , \qquad (6.1)$$

  - Based on a few given seeds by users, we can obtain a list of training samples by ranking with the relationship values.
  - As the relationship values are different, the training samples are associated with different confidence degrees, i.e. the soft label.

Soft Label SVM

- Let us consider the training data

$$(\mathbf{x_1}, s_1), \ldots, (\mathbf{x_l}, s_l) \in \mathcal{X} \times S, \quad S \subseteq [-1, +1]$$

where *s* is the soft label, the corresponding hard label set Y is obtained

$$Y = sgn(S) = \{+1, -1\}$$

The objective function is

$$\begin{aligned} \min_{\mathbf{w}, \mathbf{x} \in \mathbb{R}^m, b \in \mathbb{R}} \quad & \frac{1}{2}\|\mathbf{w}\|^2 - \nu\rho + \frac{1}{l}\sum_i y_i s_i \xi_i \\ \text{subject to} \quad & y_i((\Phi(\mathbf{x_i}) \cdot \mathbf{w}) + b) \geq y_i s_i \rho - \xi_i \ , \\ & \xi_i \geq 0, \quad i = 1, \ldots, l \ , \\ & 0 \leq \nu \leq 1, \quad \rho \geq 0 \ . \end{aligned}$$

Soft Label SVM

- The optimization problem can be solved as

$$L(\mathbf{w}, \xi, b, \rho, \alpha, \beta, \delta) = \frac{1}{2}\|\mathbf{w}\|^2 - \nu\rho + \frac{1}{l}\sum_i y_i s_i \xi_i$$
$$-\sum_i \left(\alpha_i (y_i(\Phi(\mathbf{x_i}) \cdot \mathbf{w} + b) - y_i s_i \rho + \xi_i) - \beta_i \xi_i\right) - \delta\rho \;. \qquad (6.5)$$

- By taking derivates,

$$\begin{aligned}
\frac{\partial L}{\partial \mathbf{w}} &= \mathbf{w} - \sum_{i=1}^{l} \alpha_i y_i \Phi(\mathbf{x_i}) = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^{l} \alpha_i y_i \Phi(\mathbf{x_i}) \;; \\
\frac{\partial L}{\partial \xi_i} &= y_i s_i \frac{1}{l} - \alpha_i - \beta_i = 0 \Rightarrow 0 \le \alpha_i \le y_i s_i \frac{1}{l} \;; \\
\frac{\partial L}{\partial b} &= -\sum_{i=1}^{l} \alpha_i y_i = 0 \Rightarrow \sum_{i=1}^{l} \alpha_i y_i = 0 \;; \\
\frac{\partial L}{\partial \rho} &= -\nu + \sum_{i=1}^{l} \alpha_i y_i s_i - \delta = 0 \Rightarrow \sum_{i=1}^{l} \alpha_i y_i s_i - \delta = \nu \;.
\end{aligned}$$

- The dual optimization problem

$$\min_{\alpha} \quad \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j k(\mathbf{x_i}, \mathbf{x_j})$$

$$\text{subject to} \quad \sum_i \alpha_i y_i = 0 \, , \tag{6.6}$$

$$0 \le \alpha_i \le y_i s_i \frac{1}{l}, \quad i = 1, 2, \ldots, l \, , \tag{6.7}$$

$$\sum_i \alpha_i y_i s_i \ge \nu \, . \tag{6.8}$$

- The constraint of optimization is different from regular SVM:
  - Regular SVM $\quad 0 \le \alpha_i \le \frac{1}{l}$
  - Soft Label SVM $\quad 0 \le \alpha_i \le y_i s_i \frac{1}{l}$

LRF algorithm by SLSVM

- The LRF algorithm
  - Computing the soft labels of the training data **x** corresponding to the *i-th* seed

$$s_{\mathbf{x}} = \begin{cases} R_{ix}/R_{max}, & R_{ix} \geq 0 \\ -R_{ix}/R_{min}, & R_{ix} < 0 \end{cases}$$

R_{max} Maximum of relationship

R_{min} Minimum of relationship

  - Training the data with SLSVM
  - Ranking results by the decision function of the SLSVM

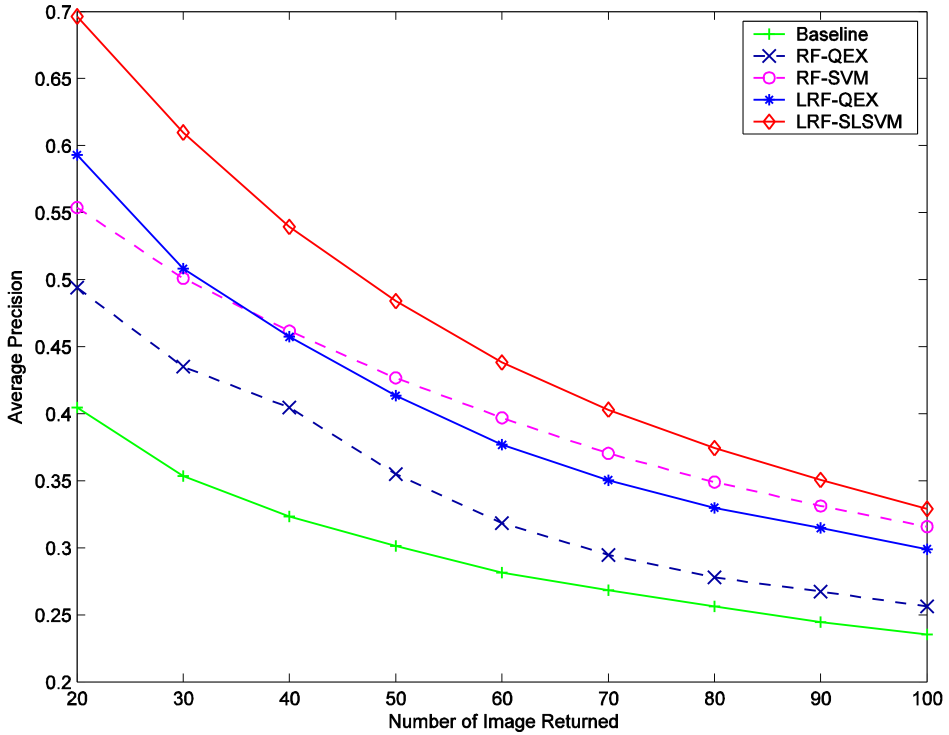
$$f(\mathbf{x}) = \sum_i \alpha_i y_i k(\mathbf{x}, \mathbf{x_i}) + b$$

Experiments

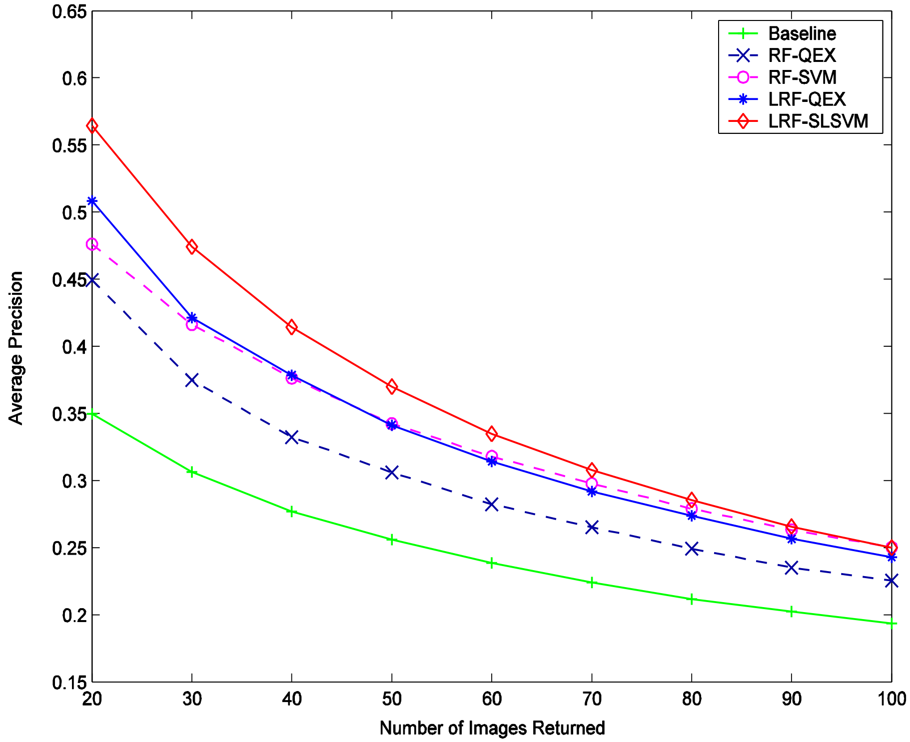
- Datasets
  - 20-Cat and 50-Cat from COREL image CDs
- Image Representation
  - Color Moment (9-dimension)
  - Edge Direction Histogram (18-dimension)
  - Wavelet Texture (9-dimension)
- Experimental Setup
  - A Log Session (LS) is defined as a basic log unit. 20 images are evaluated in each LS.
  - Schemes for comparison
    - Baseline (Euclidean distance measure)
    - Relevance Feedback – Query Expansion (RF-QEX)
    - Relevance Feedback – SVM (RF-SVM)
    - Log-based Relevance Feedback – Query Expansion (LRF-QEX)
    - Log-based Relevance Feedback – Soft Label SVM (LRF-SLSVM)

Experiments (cont.)

- For only one round relevance feedback



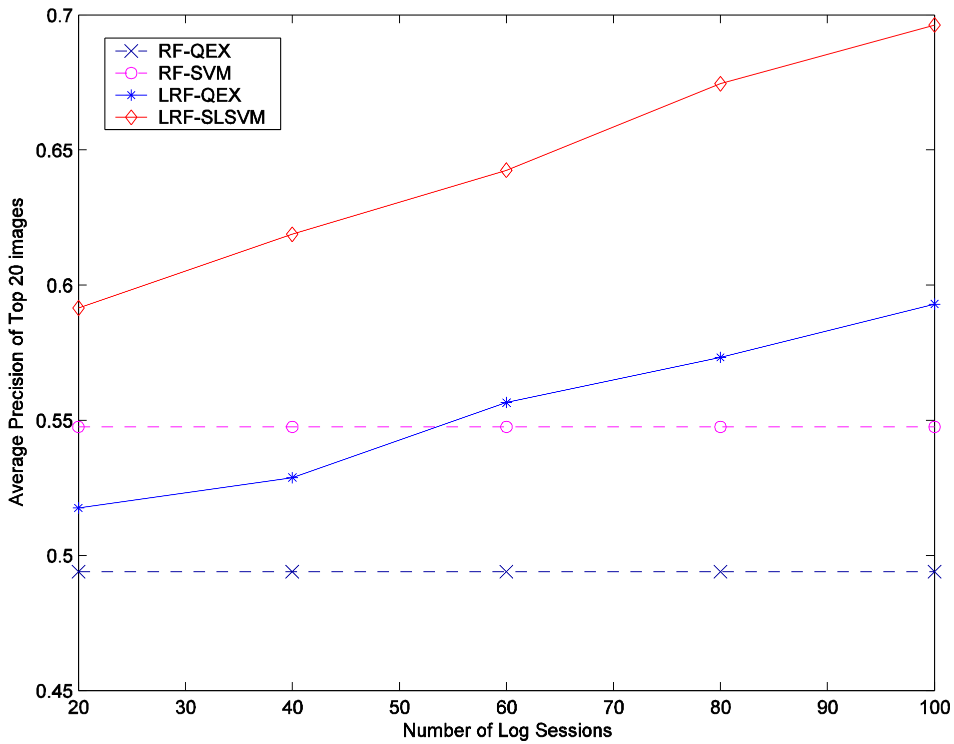
20-Cat dataset

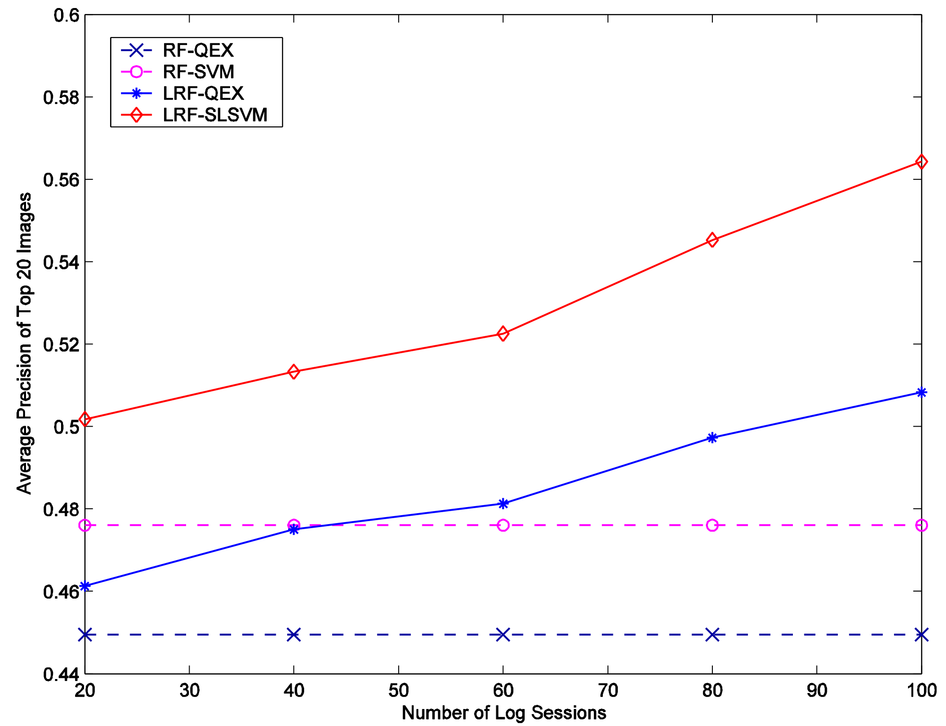


50-Cat dataset

Experiments (cont.)

- Evaluate the performance of different number of Log sessions





Experiments (cont.)

- For kernels

Table 6.1: Retrieval performance of different kernels on 20-Cat dataset

Kernel	Top 20	Top 50	Top 100
Linear	0.573 ± 0.026	0.379 ± 0.020	0.253 ± 0.014
Poly-2	0.604 ± 0.025	0.387 ± 0.020	0.257 ± 0.014
Poly-4	0.612 ± 0.025	0.397 ± 0.020	0.262 ± 0.014
RBF	$\mathbf{0.700 \pm 0.022}$	$\mathbf{0.483 \pm 0.019}$	$\mathbf{0.334 \pm 0.014}$
Sigmoid	0.597 ± 0.025	0.359 ± 0.020	0.224 ± 0.014

Table 6.2: Retrieval performance of different kernels on 50-Cat dataset

Kernel	Top 20	Top 50	Top 100
Linear	0.370 ± 0.023	0.208 ± 0.014	0.126 ± 0.009
Poly-2	0.381 ± 0.022	0.210 ± 0.014	0.130 ± 0.009
Poly-4	0.383 ± 0.023	0.212 ± 0.014	0.133 ± 0.010
RBF	$\mathbf{0.574 \pm 0.022}$	$\mathbf{0.388 \pm 0.018}$	$\mathbf{0.267 \pm 0.013}$
Sigmoid	0.422 ± 0.022	0.212 ± 0.014	0.120 ± 0.009

Outline

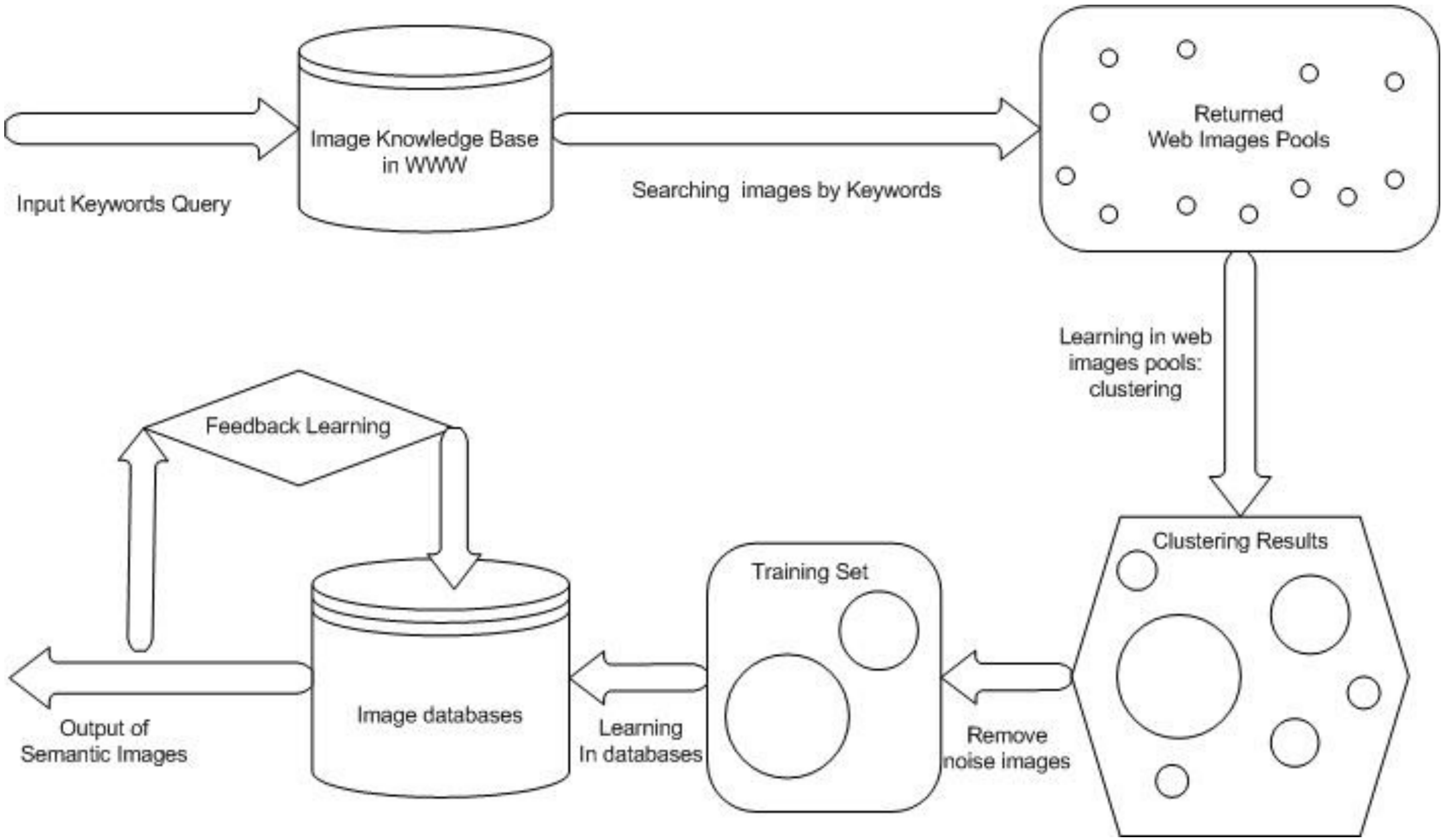
- 1. Introduction
- 2. Background & Related Work
- 3. Relevance Feedback with Biased SVM
- 4. Optimizing Learning with SVM Constraint
- 5. Group-based Relevance Feedback
- 6. Log-based Relevance Feedback
- 7. An Application: Web Image Learning
- 8. Discussions
- 9. Conclusions

7. An Application: Web Image Learning

- Motivation
  - Searching semantic concepts in image databases is an important and challenging work. Without a knowledge base, semantic understanding by computers is almost impossible nowadays.
  - Toward semantic concepts understanding, we propose to employ Web images to help on searching semantic concepts in image databases.
  - The Web images associated with keywords can served as an available knowledge base which helps the semantic learning work.
  - In order to facilitate the learning work, we suggest to engage relevance feedback with the SVMs techniques in the learning tasks.

Web Image Learning Scheme

- Proposed Architecture



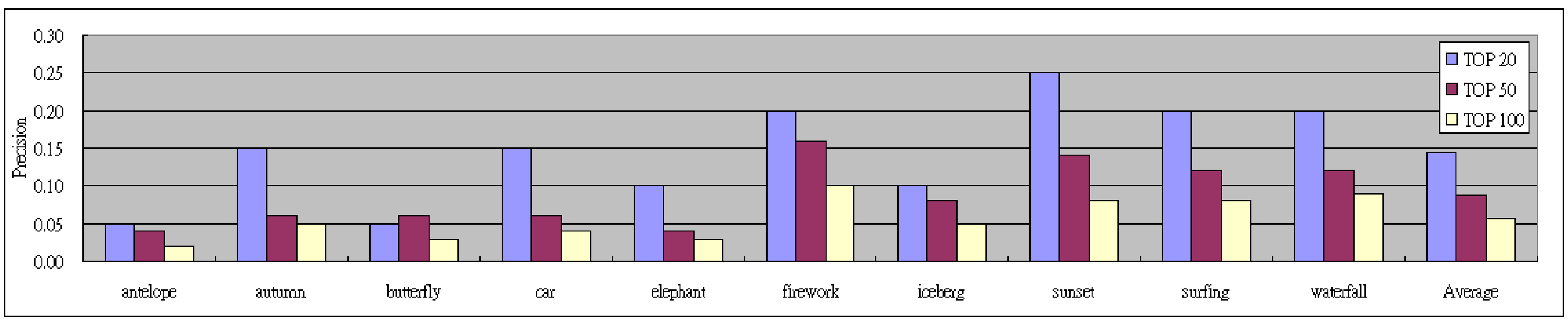
- Steps for Learning Semantic Concepts
  - Searching and clustering Web images
  - Users typing the keywords to describe the desired semantic concepts
  - Searching related Web images associated with the keywords from WWW
  - Clustering the searching results by the *k*-means algorithm
  - Removing the noisy images to obtain the final training sets of web images
- Learning semantic concepts by relevance feedback by SVMs
  - SVM provides good generalization and very excellent performance on pattern classification problems.
  - Preliminary Learning: employing one-class SVMs since only positive training samples are available.
  - Relevance Feedback Learning: engaging Biased SVMs for learning iteratively.

Experiments

- Dataset
  - Our image database contains 20,000 images selected from COREL image CDs. It includes 200 semantic categories, such as antelope, cars, and sunset, etc.
- Features
  - 9-dimensional Color Moment
  - 18-dimensional Edge Direction Histogram
  - 9-dimensional DWT texture (DB-4 wavelet, 3-level DWT)
- Experimental Setting
  - Clustering: k-means, k = 12
  - Relevance Feedback by SVMs: RBF kernel

Experiments (cont.)

- Testing semantic concepts
  - antelope, autumn, butterfly, cars, elephant, firework, iceberg, sunset, surfing, and waterfall
- Experimental results
  - Preliminary results



Experiments (cont.)

- Example: Visual experimental results for searching "firework"

Experiments (cont.)

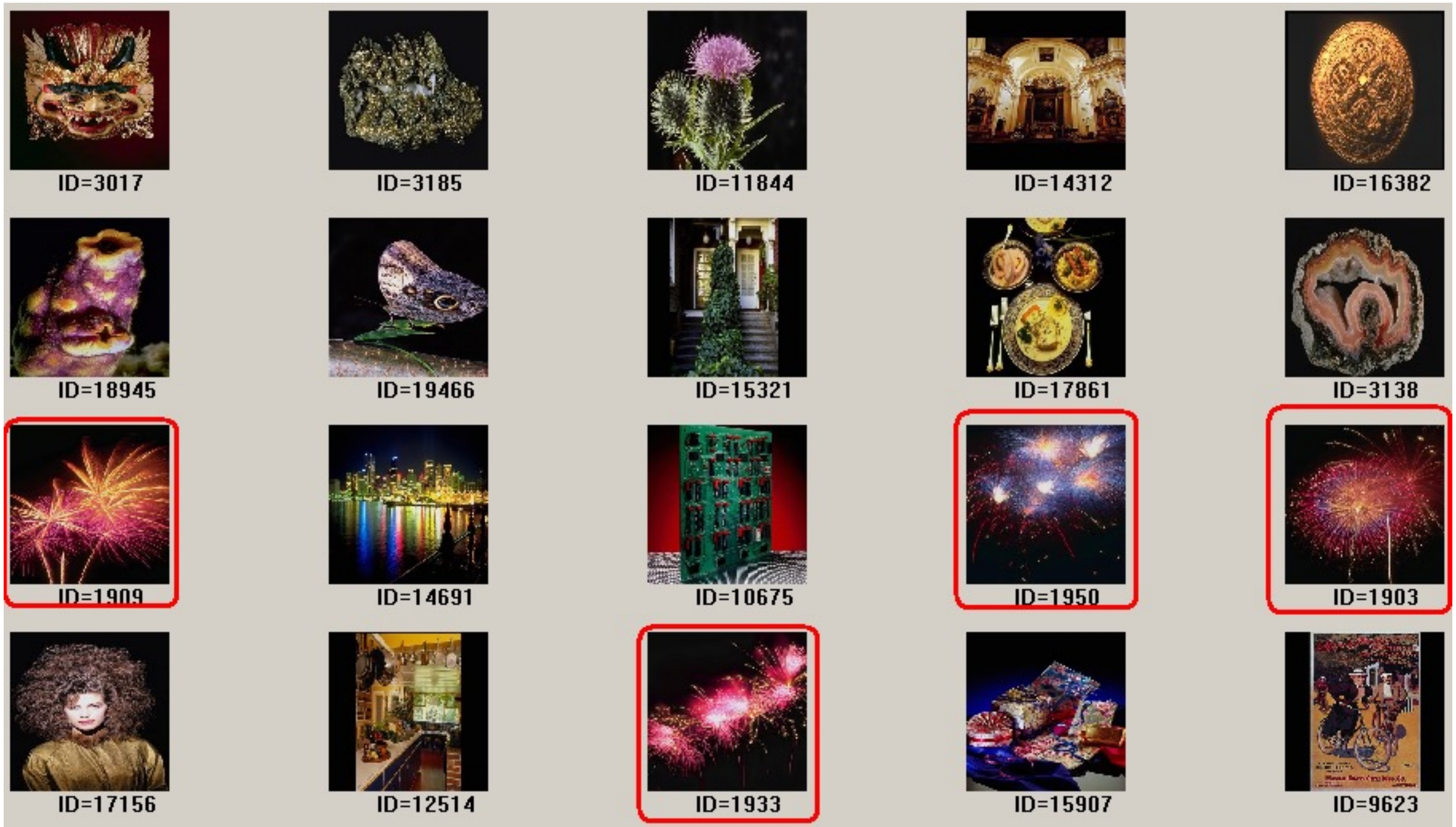
- *k*-means algorithm, *k*=12 clusters
  p=2 clusters with most samples are selected

Cluster#1

Cluster#2

Experiments (cont.)

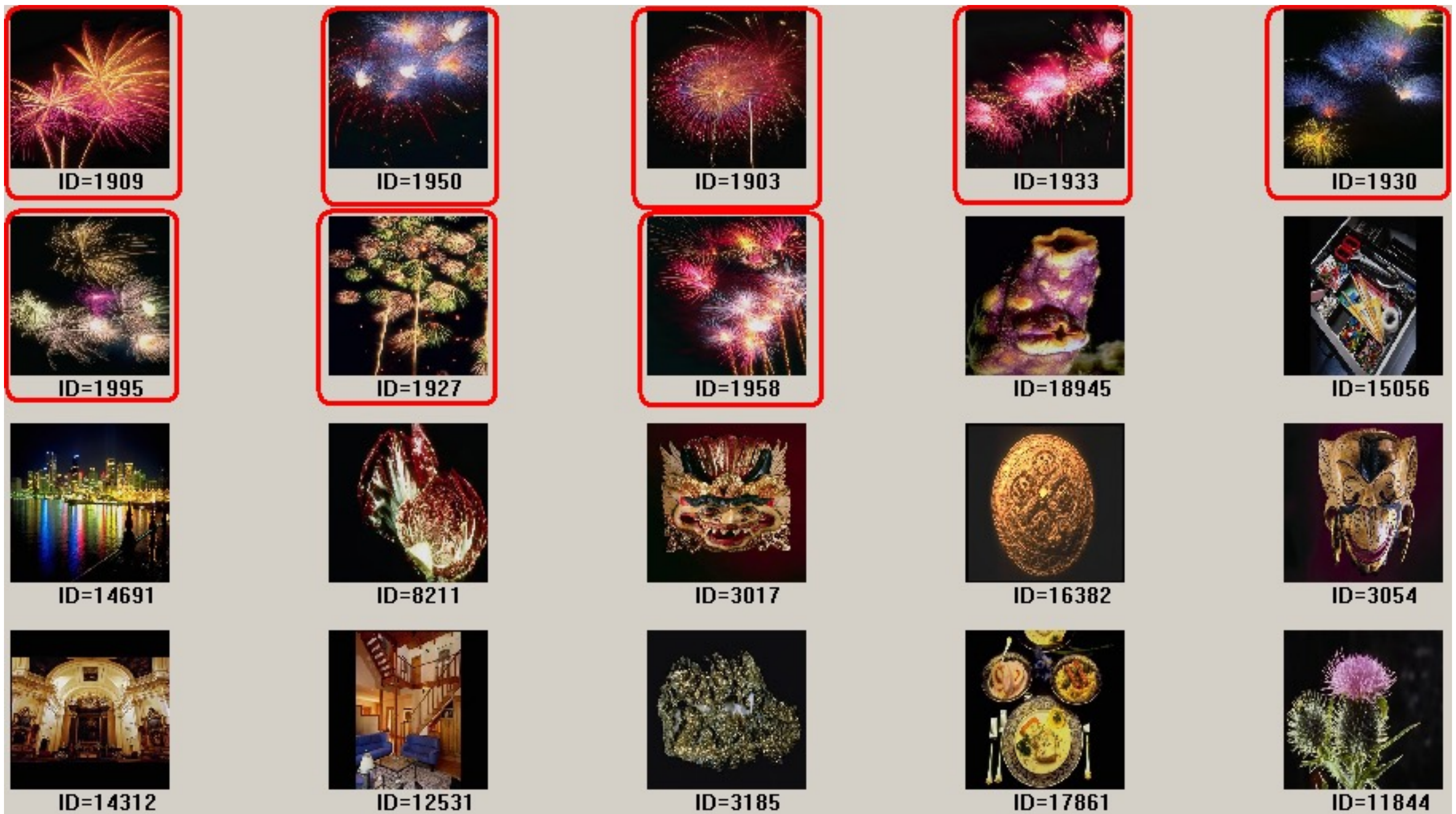
- Preliminary retrieval results from 20000 image databases



Preliminary results-Top 20

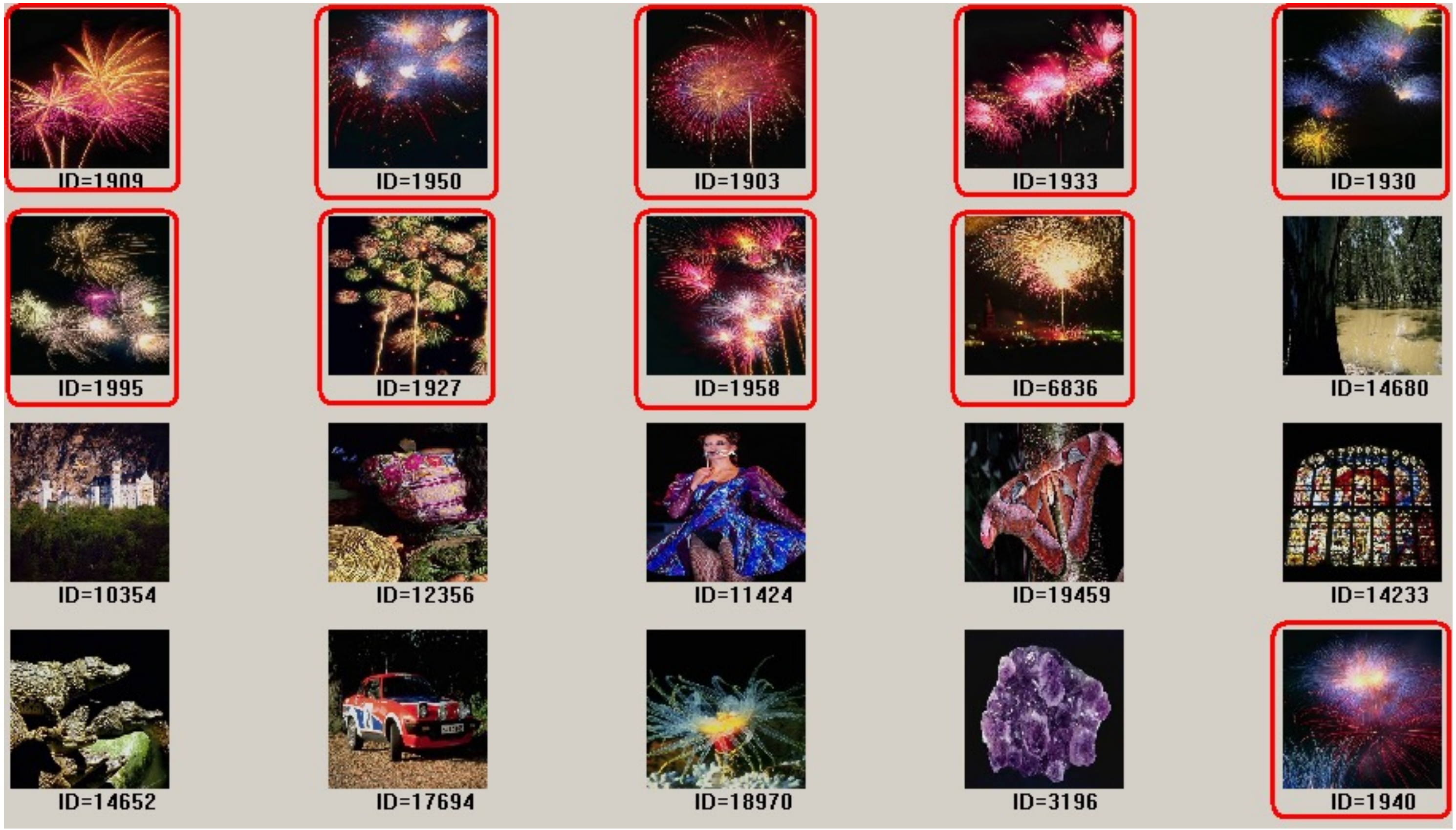
Experiments (cont.)

- learning results for relevance feedback learning



Top 20 of the 1st round Feedback results

Experiments (cont.)



Top 20 of the 2nd round Feedback results

Experiments (cont.)

Top 20 of the 3rd round Feedback results

Experiments (cont.)

- Average experimental results for relevance feedback

Table 7.1: Average retrieval precision by relevance feedback

Feedback Round	TOP 20	TOP 50	TOP 100
No Feedback	14.5%	8.8%	5.7%
1 Feedback	29.0%	15.2%	15.4%
2 Feedback	47.0%	26.4%	16.1%
3 Feedback	58.5%	32.2%	18.3%

8. Discussions

Although we have contributed much effort to studying the relevance feedback problems, limitation of our work should also be addressed.

- Limitation of our work
  - Most of our algorithms focused on the retrieval performance, but paid less attention to evaluate the efficiency problems.
  - Our proposed algorithms are based on supervised learning techniques without using the unlabeled data.
- Future Directions
  - The efficiency problems may be critical if the relevance feedback algorithms are applied in large database applications. Hence, we will consider to evaluate more detailed on the efficiency problem of our algorithms in the future.
  - Recently, semi-supervised learning techniques arouse much interest by researchers in the machine learning community. We expect these techniques could also be promising for attacking the relevance feedback problem of multimedia retrieval. However, engaging unlabeled data is a challenging work for many reliability and efficiency problems.

Outline

- 1. Introduction
- 2. Background & Related Work
- 3. Relevance Feedback with Biased SVM
- 4. Optimizing Learning with SVM Constraint
- 5. Group-based Relevance Feedback
- 6. Log-based Relevance Feedback
- 7. An Application: Web Image Learning
- 8. Discussions
- 9. Conclusions

9. Conclusions

- In this presentation, we studied the problems of relevance feedback in the context of CBIR and proposed effective algorithms to attack the learning issues.
- First, we addressed the imbalance problem of relevance feedback and proposed a Biased SVM technique to formulate the relevance feedback algorithm.
- Second, we studied two important techniques for relevance feedback and unified these two techniques for learning the similarity measure in CBIR.

9. Conclusions (cont.)

- Furthermore, we suggested to consider the data of relevance feedback as an (x+1)-class model and proposed a group-based relevance feedback algorithm using the SVM ensembles technique.
- In addition to regular relevance feedback techniques, we also studied the learning technique to improve the relevance feedback with user feedback logs. We proposed an effective SVM algorithm to attack the learning problem.
- Finally, we presented a novel and meaningful application to study Web images for searching semantic concepts in image databases. We employ a relevance feedback mechanism to attack the learning task based on SVMs techniques.

Q&A

Thank you!

Appendix

Notations for OPL and SVM

(Back)

Notation	Definition
N	number of training sample vectors
M	number of feature components (e.g. color, shape and texture components)
S	set of support vectors for a SVM classifier
L_i	dimension of the i-th feature component
$\vec{x}_n = [\vec{x}_{n1}, \ldots, \vec{x}_{nM}]$	the n-th sample vector in the image database
$\vec{x}_{ni} = [x_{ni1}, \ldots, x_{niL_i}]$	the i-th component of the n-th sample vector
$\vec{q} = [\vec{q_1}, \ldots, \vec{q_i}, \ldots, \vec{q_M}]$	the query vector
$\vec{q_i} = [\vec{q}_{i1}, \ldots, \vec{q}_{iL_i}]$	the i-th feature component of the query vector
$\vec{u} = [u_1, \ldots, u_M]$	weights of feature components
$\vec{v} = [v_1, \ldots, v_N]$	goodness values of samples
$\mathbf{W_i} = [w_{jk}]$	real symmetric full distance matrix for distance functions
$\mathbf{C_i} = [c_{jk}]$	weighted covariance matrix of samples vectors
$K(\mathbf{x}, \mathbf{y})$	Mercer kernel function for SVM
Φ	a mapping function for SVM
$f()$	decision function for SVM
$\mathcal{D}_{SVM}()$	distance function of SVM
$\mathcal{D}_{OPL}()$	distance function of Optimizing Learning
$MaxDist$	maximum OPL distance inside the positive boundary
$\mathbf{Dis}()$	overall dissimilarity measure function