Content-based Image Retrieval Using Kernel Methods

Proposal Defense

Foong Joan Tack

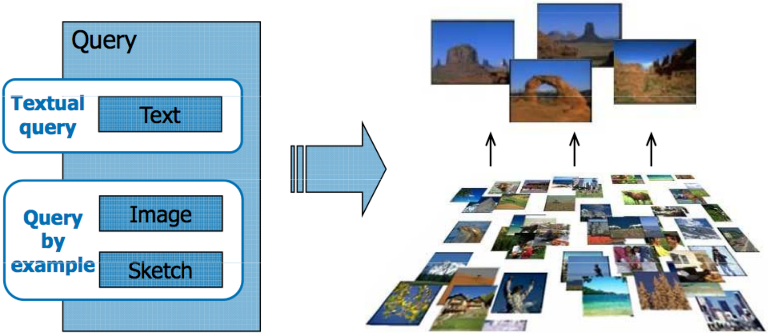
Department of Computer Science and Networked System
Faculty of Science and Technology
Sunway University
cloverevolc@yahoo.com
12058590@imail.sunway.edu.my

April 16, 2013

Overview

- What is Content-based Image Retrieval?
- What is Kernel Methods?
- Why is it a good idea to use Kernel Methods?

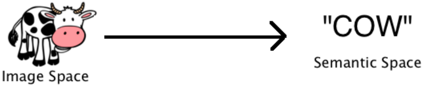
What is Image Retrieval



Courtesy of Natalia Vassilieva, Russian Summer School in Information Retrieval 2009

What is Image Retrieval

The critical problem is to determine the "meanings" of the images.



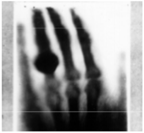
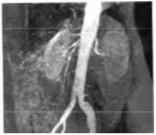
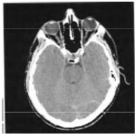
The "meanings" that we allowed to assigned to the images are to be determined by applications. In general human subject can distinguish almost 30,000 "meanings" or categories...

Application: Image Archives

- personal photo collection
- art gallery
- search for uncle John's photos
- search for Monet's painting

Application: Medical Images

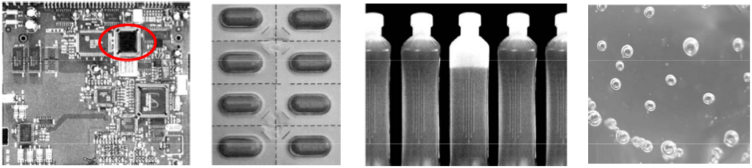
   

- x-ray
- MRI
- pathological vs healthy

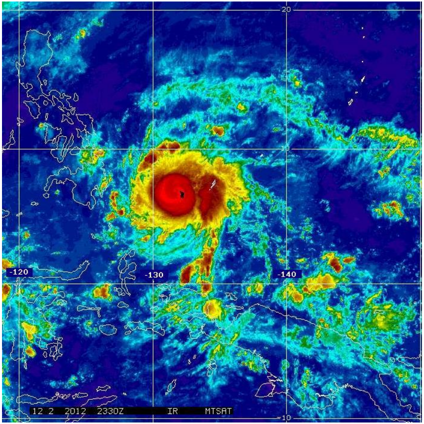
Application: Security

- suspicious items
- face recognition

Application: Industry



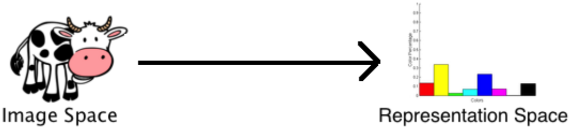
(a) CD-ROM controller (b) Pack of pills (c) Level of liquid (d) Air-bladders in plastic



- weather monitoring
- military

Representation Space

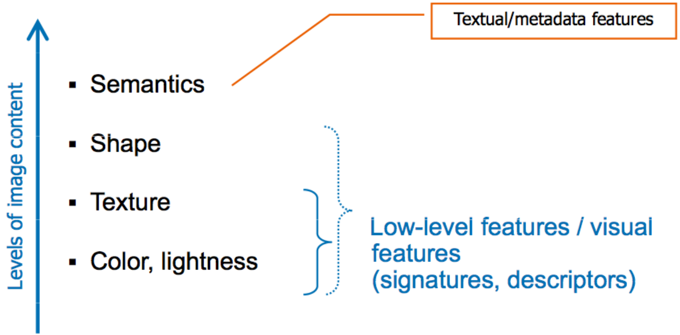
The images must be represented in a way to be processed efficiently.



Content-based vs. Description-based

	DBIR	CBIR
+	▪ Fulltext search algorithms are applicable ▪ Search results corresponds to image semantics	▪ Automatic index construction ▪ Index is objective
−	▪ Manual annotating is hardly feasible ▪ Manual annotations are subjective	▪ Semantic gap ▪ Querying by example is not convenient for a user

Features



Levels of image content
Textual/metadata features
Semantics
Shape
Texture
Color, lightness
Low-level features / visual features (signatures, descriptors)

Similarity

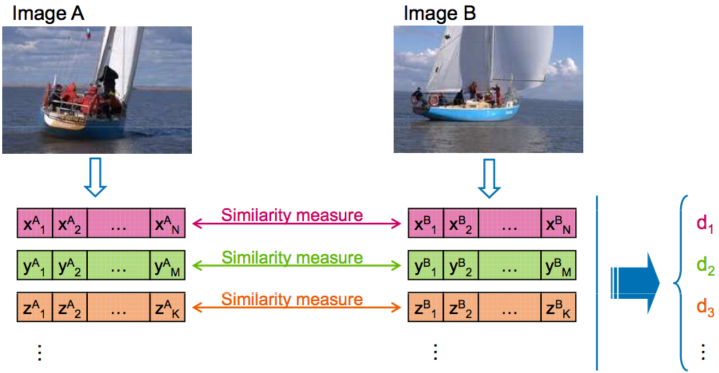
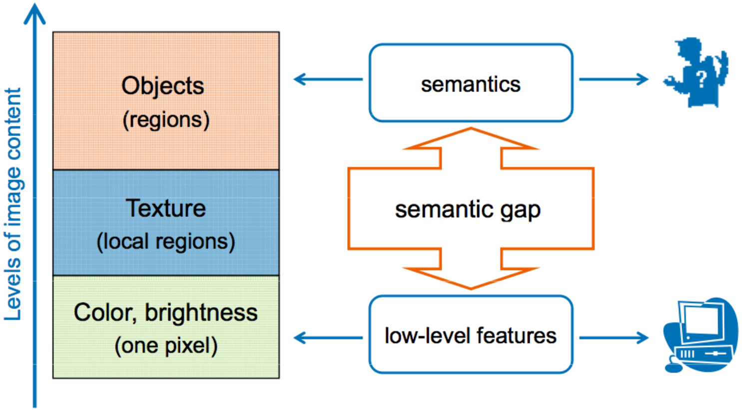
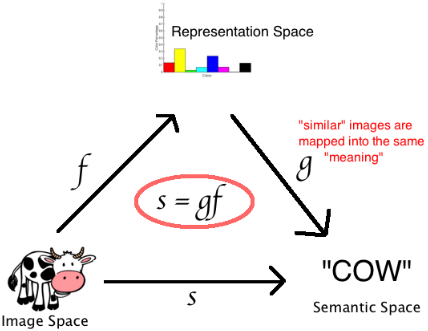


Image A
Image B
x^A_1 x^A_2 … x^A_N
Similarity measure
x^B_1 x^B_2 … x^B_N
y^A_1 y^A_2 … y^A_M
Similarity measure
y^B_1 y^B_2 … y^B_M
z^A_1 z^A_2 … z^A_K
Similarity measure
z^B_1 z^B_2 … z^B_K
d_1
d_2
d_3

Semantic Gap

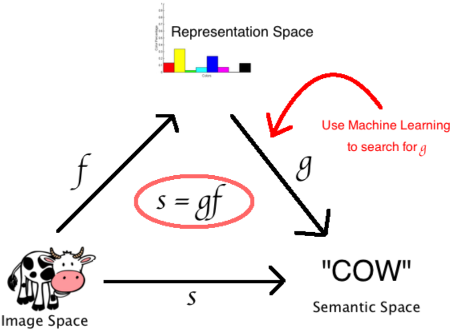


How They Fit Together?



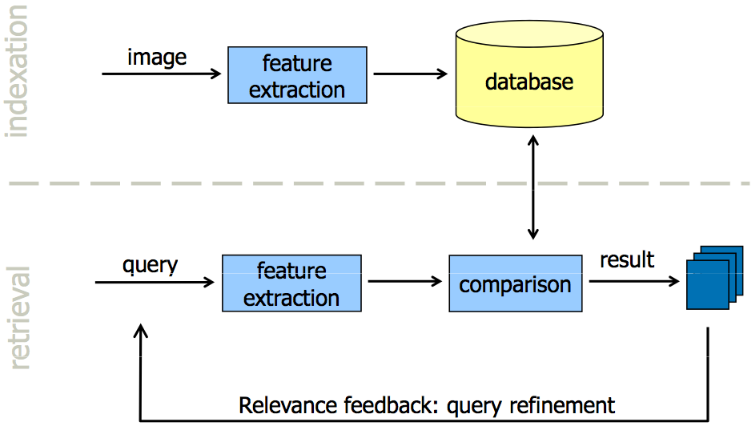
But sometime the mapping is not that simple....

Where Machine Learning Comes in?



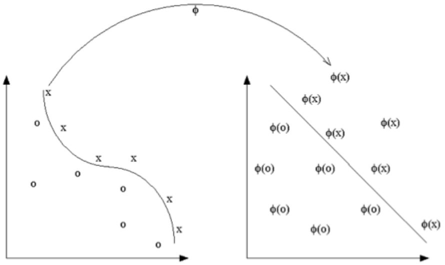
Using similarity measure is equivalent to kNN method...

Main Components



Courtesy of Natalia Vassilieva, Russian Summer School in Information Retrieval 2009

Kernel Methods



$$\phi : X \mapsto H$$

Kernel Method

We illustrate the kernel trick by a simple example. Consider the case of linear regression, we have data $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$, and wish to model it via $y = w^T \phi(x)$, where ϕ is some basis functions. The main task here is to learn w that best fit the data.

Kernel Method

The main observation here is that we can write $w = \sum a_i \phi(x_i)$, so that $y = \sum a_i \phi^T(x_i)\phi(x)$. Now the question involves only the dot product $\phi^T(x_i)\phi(x)$, which can be readily replaced by a kernel function $K(x', x)$

General Representer Theorem

Theorem

Suppose X is a non-empty set, k a kernel on $X \times X$, $\{(x_i, y_i)\}_1^n$ a training sample, g a monotonically increasing real-valued function, c an arbitrary cost function, and a class of function

$$F = \{f \in R^X | f(\cdot) = \sum_1^{\infty} \beta_i k(\cdot, z_i), \beta_i \in R, z_i \in X, |f| < \infty\}$$

Then any $f \in F$ minimizing the regularized risk functional

$$c((x_1, y_i, f(x_i), ..., (x_m, y_m, f(x_m)) + g(|f|)$$

admits a representation of the form

$$f(\cdot) = \sum_1^m \alpha_i k(\cdot, x_i)$$

Kernel Methods

Definition

A positive definite kernel $k(x, x')$ is a function from X^2 to real number such that:

$$\sum_{i,j} a_i a_j k(x_i, x_j) > 0$$

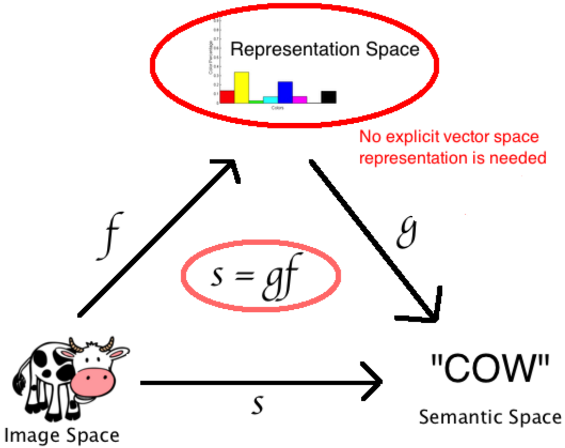
Theorem (Aronszajn theorem)

k is a positive definite kernel if and only if there exist a Hilbert Space H and a mapping $\phi : X \mapsto H$ such that $k(x, x') = \phi(x)^T \phi(x')$ for all $(x, x') \in X^2$

Kernel Methods

- no explicit representation, only kernel function
- implicitly calculate in higher dimensional space
- higher dimension $\Rightarrow$ higher capacity
- separation of data representation and algorithm
- sparte reprsentation - efficient query

Kernel Methods



Kernel Method allows no explicit representation.....just distance is enough

Definition (Kernel Trick)

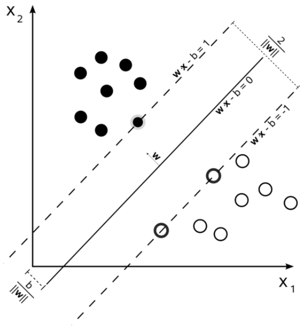
any algorithm for finite-dimensional vectors that only uses pairwise dot-products can be applied in the feature space.

The algorithm and data are separated, hence *Modularity*

Kernel Examples

- linear kernel, $k_L(x, x') = x^T x'$ and
- polynomial kernel, $k_P(x, x') = (x^T x')^d$
- Gaussian RBF kernel, $k_G(x, x') = exp(-\frac{|x-x'|^2}{2\sigma^2})$
- Given and distance function d, $k(x, y) = \frac{1}{2}(d(x, 0) + d(0, y) - d(x, y))$ is a kernel.

Support Vector Machine



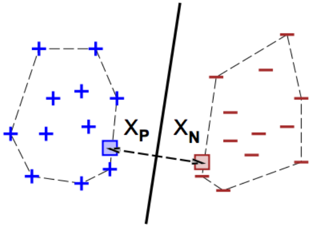
Minimize

$$\frac{1}{2}|w|^2$$

subject to

$$y_i(w \cdot x_i - b) > 1$$

Support Vector Machine - online version

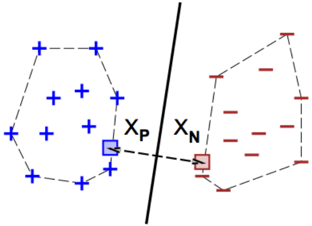


The points X_p and X_n can be parametrized as:

$$X_P = \sum_{i \in P} a_i x_i, \sum a_i = 1$$

$$X_N = \sum_{j \in N} a_j x_j, \sum a_j = 1$$

Support Vector Machine - online version



The solution can be found by optimizing:

$$max||X_P - X_N||^2$$

Support Vector Machine - online version

UPDATE(k):

- Compute $\boldsymbol{X}_P \boldsymbol{x}_k$, $\boldsymbol{X}_N \boldsymbol{x}_k$, and $\boldsymbol{x}_k \boldsymbol{x}_k$.
- Compute λ_u using equations (4) or (5).
- Compute λ using equation (6)
- $\alpha_i \leftarrow (1 - \lambda)\alpha_i$ for all i such that $y_i = y_k$.
- $\alpha_k \leftarrow \alpha_k + \lambda$.
- Update $\boldsymbol{X}_P \boldsymbol{X}_P$, $\boldsymbol{X}_N \boldsymbol{X}_P$ and $\boldsymbol{X}_N \boldsymbol{X}_N$ using equation (7) or (8).

HULLER:

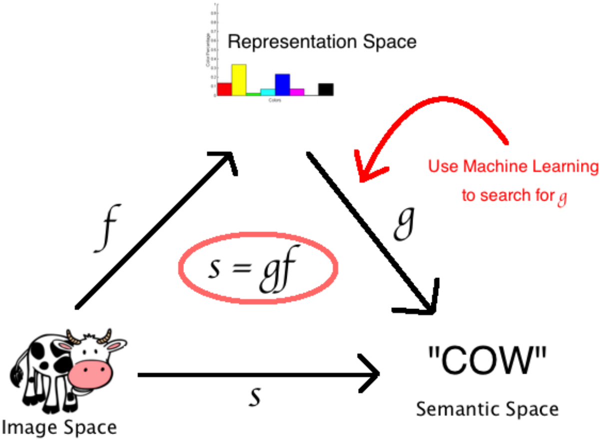
- Initialize $\boldsymbol{X}_P$ and $\boldsymbol{X}_N$ by averaging a few points.
  Compute initial $\boldsymbol{X}_P \boldsymbol{X}_P$, $\boldsymbol{X}_N \boldsymbol{X}_P$, and $\boldsymbol{X}_N \boldsymbol{X}_N$.
- Iterate:
  - Pick a random p such that $\alpha_p = 0$
  - **UPDATE**(p)
  - Pick a random r such that $\alpha_r \neq 0$
  - **UPDATE**(r)

The Huller: a simple and efficient online SVM, Antoine Bordes and Leon Bottou

Conclusion

- to side-step the need for explicit representation
- to construct a kernel as generalized distance function
- to prune the data for efficiency, for example: to use kNN to narrow down the classes label before training the SVM

Questions ?


Representation Space
Use Machine Learning
to search for g
f
g
$s = gf$
"COW"
Semantic Space
s
Image Space