France Telecom Orange Labs (Beijing) AT MSR-BING CHALLENGE ON IMAGE RETRIEVAL 2013

Lezi Wang & Shusheng Cen

wanglezi@bupt.edu.cn

Build a system that can recognize everything

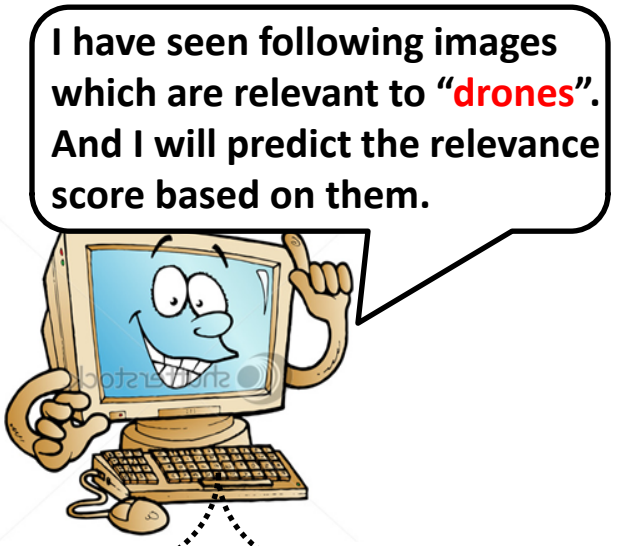
- Generic **VS** Specific information
- Both are included in the Challenge

Generic: *a man, or a person…*

Specific: *a man in a white T-shirt*,
or
David Robert Joseph Beckham
or
David Beckham…

~~Build a system that can recognize everything~~

Given a **limited** set of data, predict a reasonable order based on the relevance.



Query: Drones

The positive list

- Generate the positive list
- Compute visual similarity

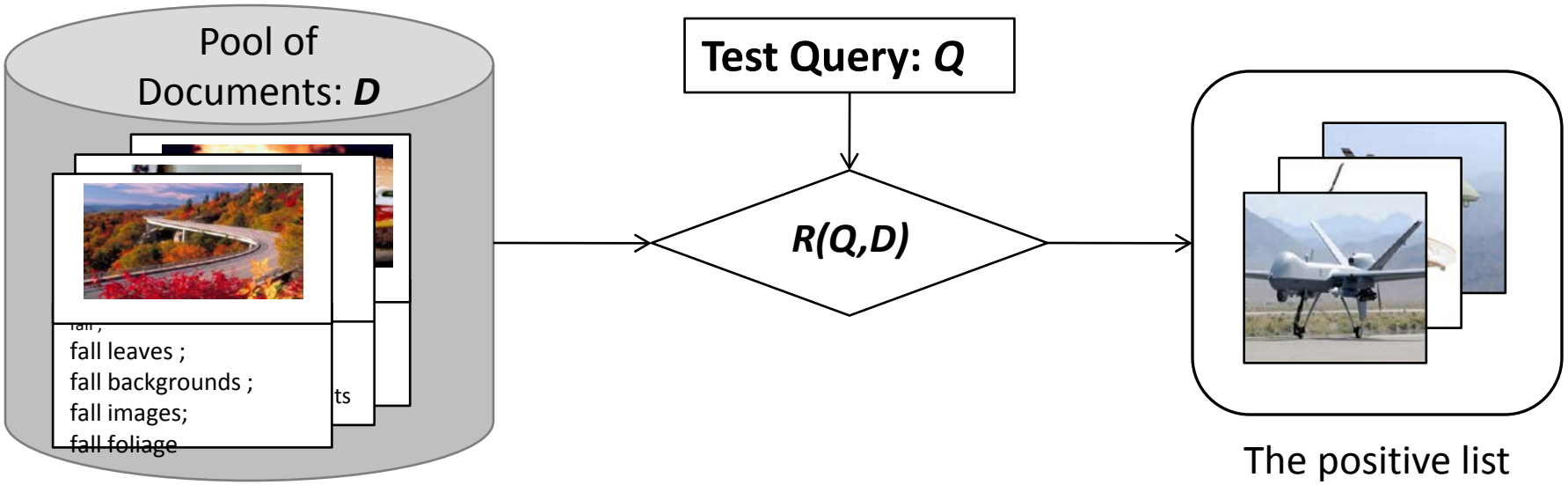
Generate positive list: Text Search

Generate documents database

- Remove “Stop Word”, Non-English logs, 23M query logs, 1M images, ~20 query terms per document on average
- Word stemming
- Query terms concatenation

assign a confidence to each query $Q_i \in D$: $Confi(D, Q_i) = \frac{C_i}{\sum_m C_m}$

Image ID	Query Term	Click
1	border colli golden retriev mix	4
1	golden retriev mix dog	3
1	border colli golden retriev mix puppi	2
1	chow chow mix	2
1	chow mix	2



- The *R(Q,D)* is defined as:

$$R(Q,D) = \sigma \sum_{j \in D} S(Q,Q_j) Confi(D,Q_j) \ \ where \ \ Q_j \in D$$

$S(Q,Q_j)$ indicates the word overlap between two queries;

$\sigma \propto click\ count\ of\ D$;

- The open source library Lucene[1] is used:

  index one query term over 23M logs < 1s

[1] *http://lucene.apache.org/*

Compute Visual Similarity: Features

Three types of features are used to describe the images:

- GIST [M. Douze, H. Jegou, et.al,2009]
- LBP-RGB Histogram(LRH) [G. Chechik, V. Sharma, et.al, 2009]
- Color-Sift(CSIFT) [G. J. Burghouts & J. M. Geusebroek,2009] + Bag of Word Model (AKM):
  - Book Size varies in 1k,10k,1M

Three runs submitted for evaluation

- Primary run

  “bootLearn’’: Cluster-based Query-Image Relevance Assessment

- Additional Runs

  “fast_v1” and “learn_RF”: no clusters, faster than primary run

Three runs submitted for evaluation

- Primary run

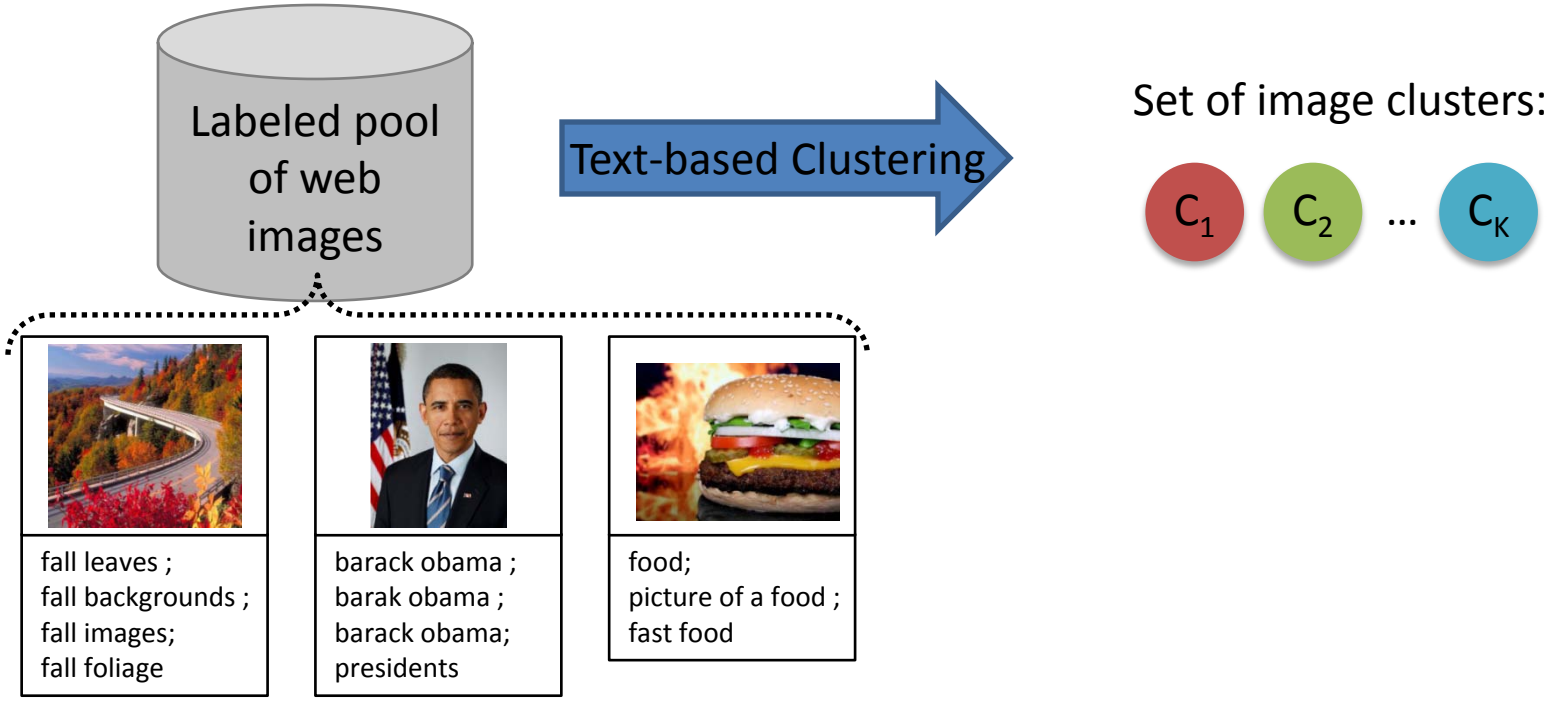
  “bootLearn’’: Cluster-based Query-Image Relevance Assessment

- Additional Runs

  “fast_v1” and “learn_RF”: no clusters, faster than primary run

Cluster-based Query-Image Relevance Assessment

- 1M images, ~10K image categories, ~200k isolated images
- A cluster can be seen as an expert judging the relevance on a query-image pair



Cluster-based Query-Image Relevance Assessment

Drones

Labeled pool of web images

Image 1
Image 2
.
.
.
Image N

Initial image list

Cluster Histogram

Select candidate clusters

Image clusters and factor α_m:

C_1 C_2 ... C_m

$\alpha_k = \beta^{rm} \times \gamma_m$

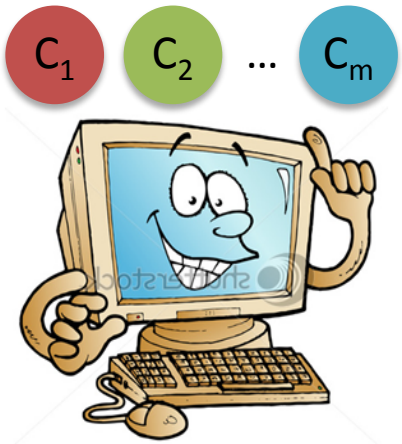
$\beta = 0.9$, r_m is the rank of C_m in the list

γ_m reflects the familiarity of the C_m to the query “drones”

Predict the relevance score

Cluster-based Query-Image Relevance Assessment

M Image clusters:



Scored

$$Rel\ (Q,I)$$

$$Rel\ (Q,I) = \sum_{m} \alpha_m \cdot S(I, C_m)$$

- *S(I, C_m)* is the average of visual similarities between *I* and top-*N* images in C_m; N =5 experimentally
- Weighted visual similarity is computed between image I_i and I_j [T.Deselaers & V. Ferrari,2011]

$$S_w(x_i, x_j) = \sum_{d} w_d(x_i^d, x_j^d)$$

Cluster-based Query-Image Relevance Assessment

- Generate positive image pairs *(x_m, x_n)* and negative pairs *(x_k, x_l)*
  - positive: *<excellent, excellent>, <excellent, good>, <pairs in same category>*

$$S_w(x_k, x_l) \le b-1 < b+1 \le S_w(x_m, x_n)$$

$$y_n(w^T X_n - b) \ge 1 \ \forall n = (i, j)$$

$$\min_{w,\xi} \frac{1}{2} \| w \|^2 + C \sum_n \xi_n$$

$$s.t. \ y_n(w^T X_n - b) \ge 1 - \xi_n \ \forall n = (i, j)$$

Cluster-based Query-Image Relevance Assessment

- ~80K query-image pairs in DEV dataset, 1000 different query terms

Method	DCG@25 (DEV)	Time Cost (one pair)
Random	0.471	-
Direct Gist	0.501	2s
weighted Gist	0.512	3s
Direct LRH	0.474	2s
weighted LRH	0.489	2s
CSIFT Booksize = 1k	0.472	5s
CSIFT Booksize = 10k	0.497	6s
CSIFT Booksize = 1M	0.529	7s

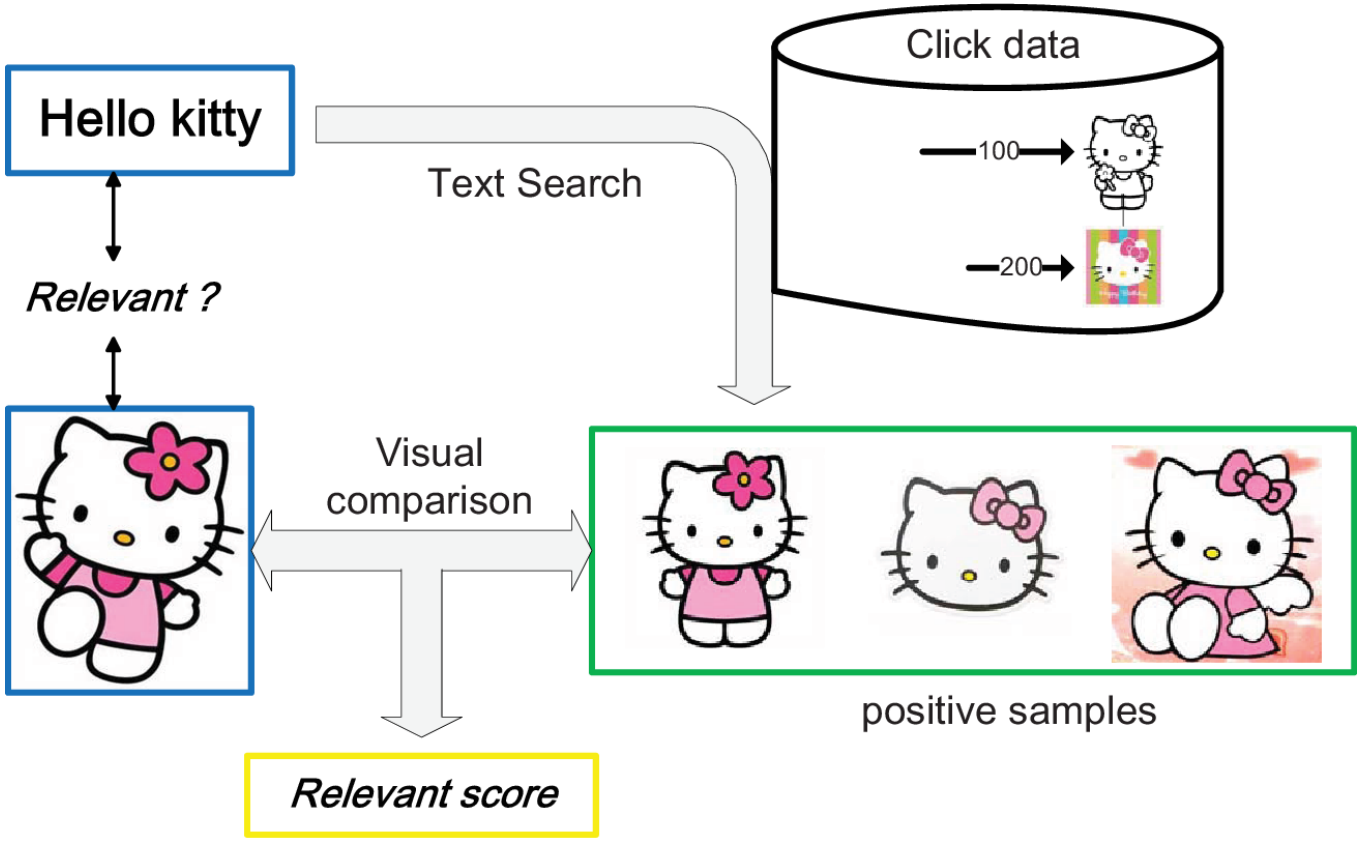
Three runs submitted for evaluation

- Primary run

  “bootLearn’’: Cluster-based Query-Image Relevance Assessment

- Additional Runs

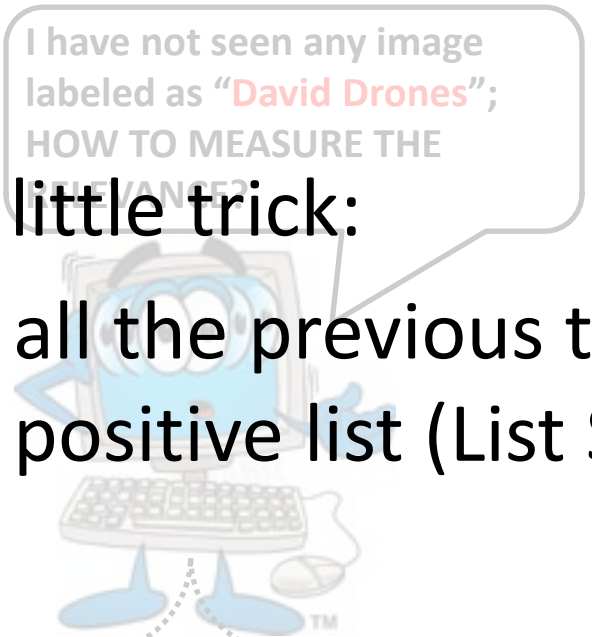
  “fast_v1” and “learn_RF”: no clusters, faster than primary run

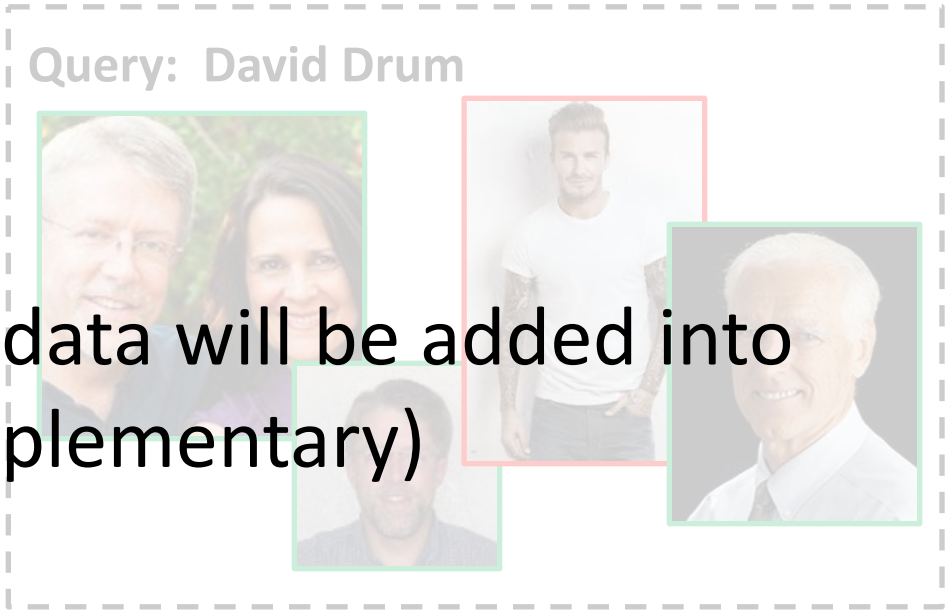

Hello kitty
Relevant ?
Text Search
Click data
100
200
Visual
comparison
positive samples
Relevant score

I have not seen any image labeled as "David Drum"; HOW TO MEASURE THE RELEVANCE?

Query: David Drum

???





A little trick:

all the previous test data will be added into positive list (List Supplementary)

???

Runs summary & Performance

- Primary run

  “bootLearn’’: Document Clustering + Text Search + Clusters Mapping + List Supplementary + CSIFT + 1M BOW

- Additional Runs:

  “fast_v1”: Text Search + Direct Gist

  “learn_RF”: Text Search + Direct Gist + List Supplementary

Runs summary & Performance

- DEV and EVA dataset are very similar in terms of the data size and content; EVA set contains 77406 query-image pairs

Run	DCG@25 on develop set	DCG@25 on test set
fast v1	0.478	0.480
learn RF	0.501	0.516
BoostLearn	0.529	0.531