A Unified Relevance Feedback Framework for Web Image Retrieval

En Cheng, Feng Jing, and Lei Zhang
Outline

- Dynamic multimodal fusion
- SRC-based textual space construction
- Friendly user interface
- Experimental results
- Conclusions
Dynamic multimodal fusion

- Image Representation
- RF in textual Space
- RF in visual Space
- Dynamic Multimodal Fusion
Image Representation

++ Image database
  📸 Photo forum sites

++ Metadata
  📸 Photographer’s comment
  📸 People’s critiques

++ Example($P_{em}$)
  • Title: early morning.
  • Category: landscape, nature, rural.
  • Comment: I found this special light one early morning in Pyrenees along the Vicdessos river near our house....
  • One of the critiques: wow...I like this picture very much...I guess the light has to do with everything...the light is great on the snow and on the sky (strange looking sky by the way)...greatly composed...nice crafted border...a beauty.
Image Representation

Textual space

Metadata

TF-IDF weighting scheme

vector space model

$F^T = \begin{bmatrix} early, \text{ morning}, \text{ landscape}, \text{ nature}, \text{ rural}, I, \text{ found}, \text{ this}, \text{ special}, \text{ light}, \text{ one}, \text{ in}, \text{ Pyrenees}, \text{ along}, \text{ the}, \text{ Vicdessos}, \text{ river}, \text{ near}, \text{ our}, \text{ house}, \text{ wow}, \text{ like}, \text{ picture}, \text{ very}, \text{ much}, \text{ guess}, \text{ has}, \text{ to}, \text{ do}, \text{ with}, \text{ everything}, \text{ is}, \text{ great}, \text{ on}, \text{ snow}, \text{ and}, \text{ sky}, \text{ strange}, \text{ looking}, \text{ by}, \text{ way}, \text{ greatly}, \text{ composed}, \text{ nice}, \text{ crafted}, \text{ border}, a, \text{ and} \text{ beauty} \end{bmatrix}$

Search Result Clustering (SRC) algorithm

Example($P_{em}$)

Query → “early morning”

$N=151, L=358, \text{ distinct terms} = 48$
Image Representation

Visually $\rightarrow$ 64-dimensional feature

Three features

- 6-dimensional color moments
  - CIE-LUV color space
- 44-dimensional banded auto-correlogram
  - HSV color space with inhomogeneous quantization into 44 colors
- 14-dimensional color texture moments
  - Fourier transform
  - CIE-LUV color space

Gaussian normalization

Normalized to $[0, 1]$
RF in textual Space and visual Space

Rocchio’s algorithm (1960)

- Optimal query
- Cosine similarity
  - Image and the initial query
  - Only relevant images
  - $\alpha=1$, $\beta=0$

Rui’s algorithm

- Clicked images
  - Optimal query
  - Mean

Feature weights
  - Standard deviation
Dynamic Multimodal Fusion

- Textual features
  - Semantic-oriented and efficient
- Visual features
  - Finer descriptive granularity
- Different feature spaces
  - Linear combination
Dynamic Multimodal Fusion

- Query
- RF in Textural Space
  - Image List
  - Non-top Images
  - Filter
- RF in Visual Space
  - User Clicked Images
  - Top Images
  - Re-ranked Top Images
- RF Fusion
- Combine
- Final Image List

Mathematical equations:

\[ S = \beta \cdot S^V + (1 - \beta) S^T \]
\[ \beta = \alpha \cdot \exp(-\lambda \cdot D_{ave}) \]
\[ D_{ave} = \frac{\sum_{i=1}^{n} \| F^V_i - F^V_{opt} \|}{n} \]
\[ F^V_{opt} = \frac{\sum_{i=1}^{n} F^V_i}{n} \]
\[ S^V = 1 - D^V \]

Fig. 1. Flowchart of the RF of the unified framework.
Search result clustering (SRC) algorithm
- Salient phrase
- Accurate and low-dimensional textual space
- Query and the ranked list of search results
  - Titles and snippets
    - Phrases ($w$)

Phrases
- Phrase frequency/inverted document frequency (TFIDF) $\text{TFIDF} = f(w) \cdot \log \frac{N}{|D(w)|}$
- Phrase length (LEN) $\text{LEN} = n$
- Intra-cluster similarity (ICS)
- Cluster entropy (CE)
- Phrase independence (IND)

$$\text{IND} = \frac{\text{IND}_l + \text{IND}_r}{2}$$
$$\text{IND}_l = - \sum_{t \in \text{l}(w)} \frac{f(t)}{\text{TF}} \log \frac{f(t)}{\text{TF}}$$
$$\text{IND}_r = \sum_{t \notin \text{l}(w)} \frac{f(t)}{\text{TF}} \log \frac{f(t)}{\text{TF}}$$

$$\text{ICS} = \frac{1}{|D(w)|} \sum_{d_i \in D(w)} \cos(d_i, c)$$

$$c = \frac{1}{|D(w)|} \sum_{d_i \in D(w)} d_i$$

$$\text{CE} = - \sum_t \frac{|D(w) \cap D(t)|}{|D(w)|} \log \frac{|D(w) \cap D(t)|}{|D(w)|}$$
**SRC-based textual space construction**

\[ x = (\text{TFIDE}, \text{LEN}, \text{ICS}, \text{CE}, \text{IND}) \]

**Linear regression model**

\[ y = b_0 + \sum_{j=1}^{p} b_j x_j + e \]

- Salient phrase
  - Salience score \( y \)
  - Top-ranke
Friendly user interface

Fig. 3. Main page of MindTracer.
Friendly user interface

Fig. 4. Detailed page of MindTracer.
Experimental results

- Evaluation dataset
- Evaluation of RF fusion
- Evaluation of SRC-based RF
- Efficiency of TVRF
Evaluation dataset

Three million images

<table>
<thead>
<tr>
<th>Query</th>
<th>Key terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eagle (3809)</td>
<td>creek, eyes, people, place, plant, sunrise, sunset, valley, bald, beautiful, fishing, flying, gray, landing, sea, shout, young</td>
</tr>
<tr>
<td>Eiffel tower (1517)</td>
<td>base, diamond, light, moon, Paris, river, sky, sunset, tier, dark, first, flying, glowing, gothic, into, middle, night, rainy, red, sparkling, top, typical, underside, up</td>
</tr>
<tr>
<td>Forest house (572)</td>
<td>animal, boat, bridge, flower, nature, lake, people, plant, street, style, autumn, snow</td>
</tr>
<tr>
<td>Greek (2646)</td>
<td>beach, building, church, coffee, farmer, goddess, island, light, man, nature, sculpture, ship, street, style, woman, sunset, white</td>
</tr>
<tr>
<td>Jaguar (337)</td>
<td>logo, people, racing, type, abstract, classic, e type, old, wild, x-type, animal, cat, cat</td>
</tr>
<tr>
<td>Merry Christmas (5266)</td>
<td>candle, card, children, gift, light, music, night, ornamenti, Santa Claus, snow, tree, red, sparkling, white</td>
</tr>
<tr>
<td>Pear (813)</td>
<td>animal, apple, blossom, leaf, shadow, tree, inside, pair, red</td>
</tr>
<tr>
<td>Rainbow (5376)</td>
<td>animal, beach, bird, bridge, falls, horse, light, lorikeet, people, plant, reflection, storm, sunset, valley, double, full, under</td>
</tr>
<tr>
<td>Tiger (3826)</td>
<td>butterfly, cat, cub, eye, flower, lily, people, Amoer, blue, common, dark, drinking, plain, Siberian, sitting, sleeping, small, Sumatran, swimming, white, yarning, young</td>
</tr>
<tr>
<td>Tulip (3743)</td>
<td>bud, field, people, proportion, blossom, colorful, dry, inside, pink, purple, red, white, yellow</td>
</tr>
</tbody>
</table>
Evaluation of RF fusion

RF fusion strategy (TVRF)

- Three parameters
  - $\alpha$ controls the overall contribution of RF in visual space
  - $\lambda$ fine-tunes the contribution
  - Scope $K$
Fig. 6. Performance of TVRF under different $\alpha$ and $K$.

Fig. 7. Performance of TVRF under different $\alpha$ and $\lambda$.

Fig. 8. Performance of the four strategies.
Evaluation of SRC-based RF

Fig. 9. Precision comparison of two RF strategies.

Fig. 10. Efficiency comparison of two RF strategies.
Efficiency of TVRF

Fig. 11. Efficiency comparison of TVRF and SRC-TVRF.
Conclusions

- Textual feature-based RF (TBRF) and visual feature-based RF (VBRF)
  - Quickly select a possibly relevant image set
  - Further re-rank
- Textual feature-based RF mechanism
  - Search result clustering (SRC) algorithm
  - Accurate and low-dimensional textual space
- UI
  - Implicit relevance feedback
Thanks for attention

Thanks for attention

Thanks for attention

Thanks for attention