Trans Media Pseudo Relevance Feedback

J. Ah-Pine, C. Cifarelli, S. Clinchant, G. Csurka, and J. Renders. XRCE’s Performance evaluation over training images among visual neighbors [2]:

Trans Media Relevance Feedback (TMRF) defines a multimodal similarity using visual and textual similarities.

- We generalize TMRF [1] to obtain a probabilistic model.
- We extend TagProp [2], by the integration TMRF.
- Even though the new image contains only visual information, we improve over visual-only auto-annotation (TP).

Image Annotation with Tagprop

- \( p_{w,u} \) denotes absence/presence of word \( u \) for image \( i \).
- Keyword presence probability is a weighted sum of keyword presence among visual neighbors [2]:
  \[
  p(y_{ij} = +1) = \sum_j p(y_{ij} = +1) p(j)
  \]

- Distance-based weights \( p(j) = \frac{\exp(-d_{ij})}{\sum_j \exp(-d_{ij})} \).
- Single parameter controls decrease of weights with distance.
- Linear combination of distances: \( d_{ij} \rightarrow w_i d_{ij} \).
- Optimize leave-one-out log-likelihood over training images
  \[
  L = \sum_i \sum_{w} c_{w} \log p(y_{ij})
  \]

- \( c_{w} \) is a cost to balance difference in keyword presence/absence:
  - More tag absences than presences, and
  - Absences are much noisier e.g. user gives a handful relevant tags.

Experimental Setup

- Experiments on COREL 5K and IAPR-TC 12 datasets.
- Visual Features
  - Same 15 visual features as in [2], including Gist, global color histograms, and local SIFT and Hux descriptors.
  - Often used as “Joint Equal Contribution” (JEC) distance.
- Text Features
  - Tag distance: the intersection over union of tags of two images
    \[
    d_{ij} = 1 - |Y_i \cap Y_j|/|Y_i \cup Y_j|
    \]
  - Test distance (IAPR only) based on the cross-entropy of full captions.
- Performance evaluation
  - keyword-based retrieval: mean Average Precision (MAP) and Break-Even Precision (BEP).
  - annotation-performance: image MAP (iMAP), and image BEP (iBEP).

COREL 5K Dataset

- Contains 4500 train images, 500 for test images, 1:5 annotation tags each.
- Annotation tags designed for keyword-based retrieval.
- TMRF outperforms TagProp, and LTP generally outperforms STP on this dataset.

IAPR-TC12 Dataset

- Contains about 18,000 train images and 2,000 test images, all with descriptions.
- Annotation keywords are the extracted common nouns from the captions.
- Experiments with single or multiple base distances, and \( d^2 = \{\text{Tag}, \text{Text}\} \).
- Single distance STP > LTP, while for multi distances LTP > STP.
- Improvements: 1% in iMAP, and 5% in MAP and BEP.

Table 1: Combining different \( d^2 \) distances, \( J = 400 \), and \( K = 10 \)

<table>
<thead>
<tr>
<th>( d^2 )</th>
<th>MAP</th>
<th>BEP</th>
<th>iMAP</th>
<th>iBEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>{Tag}</td>
<td>35.4</td>
<td>36.0</td>
<td>47.0</td>
<td>42.6</td>
</tr>
<tr>
<td>{Text}</td>
<td>35.6</td>
<td>36.3</td>
<td>36.3</td>
<td>47.4</td>
</tr>
<tr>
<td>{Tag, Text}</td>
<td>35.9</td>
<td>36.3</td>
<td>36.3</td>
<td>48.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( d^2 )</th>
<th>MAP</th>
<th>BEP</th>
<th>iMAP</th>
<th>iBEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>{Tag}</td>
<td>36.0</td>
<td>36.7</td>
<td>49.6</td>
<td>44.4</td>
</tr>
<tr>
<td>{Text}</td>
<td>36.4</td>
<td>36.7</td>
<td>49.6</td>
<td>44.4</td>
</tr>
<tr>
<td>{Tag, Text}</td>
<td>36.2</td>
<td>36.6</td>
<td>49.9</td>
<td>44.4</td>
</tr>
</tbody>
</table>

References
