Minimizing risk in video hyperlinking

Presented by
Chong-Wah Ngo
City University of Hong Kong

Zhi-Qi Cheng and Xiao Wu
What makes a video “link target”?

- Supplementing anchor
- Serendipity

User experience by minimizing:
- false link
- redundancy

Prefer popular and “easy” targets
A point $x$ is popular if many other points regard $x$ as "friend".

Hub score of a point $x$:

$$N_k(x) = \sum_{i=1}^{n} P_{i,k}(x)$$

$x$ is hub if $N_k(x) > k$

The **minimal number of dimensions** required to **describe** a point w.r.t to its local neighborhood.

**Easiness** – **Local Intrinsic Dimension (LID)**

**volume expansion** \(\rightarrow\) **LID** \(\rightarrow\) **Risk**

Easiness – Diversity

Average pairwise distance between a target and its $k$-nearest neighbors

safer region

risky to link highly redundant trivial information
Insights of 122 anchors on development set

- diversity increase
- diversity decrease
Insights of 122 anchors on development set
Insights of 122 anchors on development set
Insights of 122 anchors on development set

local intrinsic dimension

hubness
Insights on dataset

Intrinsic dimension of dataset: 53
LID of 122 development anchors: 33
LID of 25 testing anchors: 23.4
Algorithm – the art of compromise

Hub 👆
Local intrinsic dimension (LID) 👆
Diversity 👆

Optimization: Select $k$ out of $n$ candidate targets

$0$-$1$ assignment vector $\text{hub}$ $\text{LID}$ $\text{distance matrix}$

\[
\max_{Y} \left\{ \frac{Y^t H}{k} - \frac{Y^t D}{k} + \frac{Y^t A Y}{k(k - 1)} \right\}
\]

Solution

– Relax the $\{0,1\}$ constraint to $[0,1]$
– Similar to quadratic programming problem

On the selection of anchors and targets for video hyperlinking, in ICMR 2017
Variants of algorithm

Depending on the initialization of assignment vector $Y$

**Hub-first**
Initialize the first $k$ targets with largest hub scores to 1

**LID-first**
Initialize the first $k$ targets with largest LID scores to 1

**Intuition**
- Hub-first for anchor selection
- LID-first for target selection

Popular content
Specific content
**Submissions**

*Run-1*: Visual baseline

*Run-2*: Run-1 + LID-First (re-rank top-100)

*Run-3*: Multimodal baseline

*Run-4*: Run-3 + LID-First (re-rank top-100)
Implementation

• Exclude 2,719 testing videos without speech – intuitively not suitable as targets?
• Use LDA-based model for video fragmentation (ACL 2017)
• Visual run based on 14K concepts
  – ImageNet, ImageNet-Shuffle, SIN, RC, Places
• Use LIMSI ASR
• Multimodal run based on the fusion of cosine similarity and Siamese network
Cross-modal evaluation

Feed different input pairs
- visual, visual
- text, text
- text, visual
- visual, text

Softmax has two nodes – Probability of similarity and dissimilarity

Average fusion of pair similarities

Siamese recurrent architecture – train using 122 anchors of development set

Learning to rank question answer pairs with holographic dual LSTM architecture in SIGIR 2017
## Result

<table>
<thead>
<tr>
<th></th>
<th>P@5</th>
<th>P@10</th>
<th>P@20</th>
<th>MAP</th>
<th>MAiSP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run-1</td>
<td>0.864</td>
<td>0.852</td>
<td>0.502</td>
<td>0.1848</td>
<td>0.1113</td>
</tr>
<tr>
<td>Run-2</td>
<td>0.864</td>
<td>0.860</td>
<td>0.530</td>
<td>0.1849</td>
<td>0.1128</td>
</tr>
<tr>
<td>Run-3</td>
<td>0.856</td>
<td>0.852</td>
<td><strong>0.582</strong></td>
<td><strong>0.1951</strong></td>
<td><strong>0.1199</strong></td>
</tr>
<tr>
<td>Run-4</td>
<td>0.856</td>
<td>0.852</td>
<td>0.710</td>
<td>0.2392</td>
<td>0.1473</td>
</tr>
</tbody>
</table>

**Conclusion-1**: Multimodal run brings some improvement for search depth @ 20 and beyond
## Conclusion-2: LID-first boosts multimodal run and shows the best improvement for search depth @ 20 and beyond
Correlation between hub & performance

Hub scores of 25 testing anchors
Correlation between LID & performance

LID of 25 testing anchors
Sahaja Yoga treats drug addiction and disease

Shri Mataji started Sahaja Yoga @ India in 1970

Anchor 145
Yoga practice

shower, 0.970
shoji, 0.941
window screen, 0.456
television, television system, 0.404
ballet dancer, 0.341
dress, 0.313
home, 0.270
balance beam, beam, 0.232
Adult_Female_Human, 0.220
Speaking_To_Camera, 0.209
leotard, unitard, body suit, cat suit, 0.180
How LID-first boosts performance

Anchor 124
University marching band
How LID-first boosts performance

Anchor 124
University marching band
When does it fail?

• Name entities in ASR are recognized incorrectly
  – anchors 124, 125, 133, 135, 140, 141, 147

• Data statistics alone is insufficient
  – May pull context-irrelevant but popular and safe fragments to a higher rank
  – Example: anchors 130 (food preparation), 139 (hat show)
Conclusion

• Multimodal run diversifies link targets
• Hub + LID + diversity improves P@20, MAiSP, MAP
• Some correlation between hub+LID of anchors and performances
• More analysis is required to understand the performance …