SYSTEM OVERVIEW

A. With pre-trained Machine Learning Models

What is a good image search term and what is not? There is no real-world evidence for this, therefore training data had to be assembled from existing information.

When BBC News uses photos by Getty Images, they sometimes expose the image’s ID. Exploiting this, we downloaded >1500 articles along with image meta data from the Getty database. (see Fig. 2)

Assuming that each term in the image meta data suffices as a query to find exactly that image, we generated our training data. Article terms were matched with the image descriptions: Each term that occurred both in the article and in the image was labelled as search term.

From this information, the networks learned which feature values make a good search term.

B. With simple Statistics

Assuming that the most relevant terms occur often and early, their relevance is calculated as follows:

\[ p(t_f, f_o) = \frac{t_f}{\max(t_F)} \times (1 - f_o) \]

where \( t_f \) denotes the frequency of a term, \( f_o \) its first occurrence value and \( \max(t_F) \) the highest frequency value of all terms in the article.

Features of a Term

Time waits for no man. Unless that man is Chuck Norris.

- **Term frequency (tf)** is the number of times a term occurs in an article.
  \( tf("man") = 2 \)
- **First occurrence (fo)** is the relative position in an article at which a term occurs for the first time – 0 representing the very first character of the article and 1 the last.
  \( fo("man") = 0.3273 \)
- **Entity category** (ec) is a nominal value describing some specific groups of terms, such as “Event”, “HumanProtagonist” or “Location”.
  \( ec("Chuck Norris") = HumanProtagonist \)

PERFORMANCE

**Machine Learning vs. Statistics**

<table>
<thead>
<tr>
<th></th>
<th>ML (A) - best case</th>
<th>ML (A) - average</th>
<th>Statistical (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>factual correct</td>
<td>47</td>
<td>41</td>
<td>10</td>
</tr>
<tr>
<td>factual incorrect</td>
<td>8</td>
<td>10</td>
<td>60</td>
</tr>
<tr>
<td>no image</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

**Evaluation methodology**

- Sample of 100 BBC articles
- For each article, image selection was run with 4 neural networks + the statistical approach
- Selected images were classified manually as "factual correct" or "factual incorrect", according to our own definition of factual correctness

**Impact of the Features**

<table>
<thead>
<tr>
<th>Term Frequency</th>
<th>First Occurrence</th>
<th>Entity Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>improving</td>
<td>improving</td>
<td>improving</td>
</tr>
</tbody>
</table>

**Performance per Article Topic**

- **World**
- **UK National**
- **UK Regional**
- **International**
- **Special Interest**

**Fig. 6: Number of correct images, by article topic and term ranking approach**

**Fig. 5: Change in the number of correct images that resulted from adding one specific feature to the neural networks**

**Fig. 4: Average performance of approaches A and B, including the best performing run for comparison**

**Fig. 3: Training a neural network for term ranking**

**Fig. 2: Scanning the corpus**

**Fig. 1: Overview of the image selection pipeline**

**A. Article Preprocessor**

turns the plain text of a news article into a list of terms. Each term is described by certain features.

**B. Using these features, two different ranking mechanisms predict how relevant each term is for the image search.**

**The most relevant terms are composed into a query string.**

**This string is used to query an image database.**

**Fig. 1: Overview of the system pipeline**

**A. Key visualizations**

- Article Text ➔ Article Preprocessor ➔ Term Ranking ➔ Query Generation ➔ Image Search

**B. Key entities**

- PubMed articles
- Image meta data
- Statistical scores
- Machine learning scores

**C. Key features**

- **First occurrence**: the relative position in an article at which a term occurs for the first time – 0 representing the very first character of the article and 1 the last.
- **Term frequency**: the number of times a term occurs in an article.
- **Entity category**: a nominal value describing some specific groups of terms, such as “Event”, “HumanProtagonist” or “Location”.

**Fig. 4: Average performance of approaches A and B, including the best performing run for comparison**

**Fig. 3: Training a neural network for term ranking**

**Fig. 2: Scanning the corpus**

**Fig. 1: Overview of the system pipeline**

**A. Article Preprocessor**

turns the plain text of a news article into a list of terms. Each term is described by certain features.

**B. Using these features, two different ranking mechanisms predict how relevant each term is for the image search.**

**The most relevant terms are composed into a query string.**

**This string is used to query an image database.**

**Fig. 1: Overview of the system pipeline**