Word2Image: Towards Visual Interpretation of Words

Haojie Li
National University of Singapore

Introduction

• Motivation
  ▫ A picture is worth 1000 words
  ▫ Traditional dictionary
    ▪ Contain word entries accompanied by photos or drawings to illustrate what the words mean
  ▫ Online visual dictionaries
    ▪ Merriam Webster Visual Dictionary Online
    ▪ The Visual Dictionary Online
    ▪ Serves as educational role
    ▪ Manually strictly picked image
    ▪ Not all concepts have corresponding images
    ▪ Images are usually not real world images

• Our goal
  ▫ To develop a system that can automatically generate sets of real images to visually interpret a given word
  ▫ Leverage the community-generated online multimedia content such as Flickr

Challenges

▫ Correctness
▫ Diversity
▫ Representativeness

Approach

• Framework
Generating diverse and precise image set

- Heuristic for diversity
  - Sample images from different groups, different users, different time
- Correlation analysis for precision
  - Flickr related tags: high co-occurrence

\[ \text{CorrScore}(J, w) = \| \{ w' | w' \in RT_J \land w' \in \text{TagJ} \cup \text{TitleJ} \} \]  
- Image J is accepted as relevant if \( \text{CorrScore}(J, w) > T_h \)

Discover diversity using semantic clustering

- Computing salience score of keywords
  - Different keywords in the image's tags contribute differently to the discovery of topics
  - E.g., “trunk” dominates over “water” in finding interested topics for “elephant”
  - Statistical and knowledge-based measure
  - Flickr TF-IDF of tags
  - Wordnet: hyponyms and synonyms
  - Hyponymy: “athlete” includes “acrobatic”, “baseball player”, “tennis player”, “runner”
  - Meronymy: “tusk” and “trunk” are meronyms of “elephant”

- Top-10 salient keywords for “elephant”:
  - African, tusk, wildlife, trunk, safari, zoo, Thailand, animal, nature, India

- Text-based clustering of images

Comparison

- Before salient words discovering:
  - Zoo, animal, Africa, animals, safari, London, wildlife, Kenya, nature, Tanzania
- After salient words discovering:
  - African, tusk, wildlife, trunk, zoo, safari, Thailand, animal, nature, India

African, tusk, wildlife, trunk are more representative than the original tag set. Using saliency words discovering, we can find more meanings given a certain concept.

Clustering in Visual Space

- Correlation analysis: Flickr tags provide additional information
- Semantic Clustering: given the salience score of key words, cluster images
- Visual Clustering: discover visual consistent sub-clusters after visual clustering
Generating representative images using visual clustering

• K-means on the visual (grid of color moments) space of each semantically consistent cluster
• Ranking each sub-cluster
  • the sum of saliency score of keywords in the cluster
  • the number of images in the cluster
  • the coherence of the cluster
• Select representative image from top sub-clusters

Experimental results

• 2 types of evaluation
  • Objective evaluation on the precision
  • Subjective evaluation on the diversity & representativeness
• 25 Concepts
  • elephant, camel, buildings, athlete, pyramid, holidays, temple, flower, bridge, ...

Experimental results

• Precision evaluation
  • To validate the effectiveness of correlation analysis in improving the accuracy of retrieval and generating representative images
  • Baseline: tag-based
  • Metric:
    • the precision for image retrieval (P-IR) of icon images
    • the precision at generating top-10 (P@10) and top-20 (P@20) representative images

<table>
<thead>
<tr>
<th>Metric</th>
<th>Tag based</th>
<th>Tag+Correlation based</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-IR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P@10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P@20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Experimental results

• User study
  • To highlight the system’s usability and performances on discovering diversity and representativeness

Table of Contents

• Introduction
• Discovering Semantic and Visual Diversity
• Experiment and result
• Conclusion
Experimental results

- More examples: pyramid

- Discovered topics:
  - "France-Paris-museum-louvre"
  - "Africa-Egypt-cairo-desert"
  - "Mexico-yucatan-maya-temple"
  - "history-architecture-giza-sphinx"

Experimental results

- More examples: holidays

- Discovered topics:
  - "Winter-december-happy-xmas"
  - "Beach-sea-ocean-sun"
  - "Disneyland-disney-california-travel"
  - "Vacation-travel-hotel-happy"

Experimental results

- More examples: athlete

- Discovered topics:
  - "Run-marathon-race-track"
  - "Run-swim-ironman-bike"
  - "Soccer-girl-ball-woman"
  - "Basketball-ball-people-high"

Table of Contents

- Introduction
- Discovering Semantic and Visual Diversity
- Experiment and result
- Conclusion

Conclusions

- Introduce a framework which attempts to leverage the web image collection to translate a word into its visual counterpart
- Generate high quality, precise, diverse and representative images given a certain concept
- The preliminary experimental results have demonstrated its usability and effectiveness.

Future works

- More experiments to evaluate the use of social media for auto-visual dictionary task

- Multimedia dictionary
  - To build a large-scale multimedia dictionary, where multi-modality information including image, video, audio and text are integrated to explain concepts.
Q and A