Retrieving Diverse Social Images Task
- task overview -

Bogdan Ionescu (UPB, Romania)
Adrian Popescu (CEA LIST, France)
Mihai Lupu (TUW, Austria)
Henning Müller (HES-SO in Sierre, Switzerland)

Outline

- The Retrieving Diverse Social Images Task
- Dataset and Evaluation
- Participants
- Results
- Discussion and Perspectives
Diversity Task: Objective & Motivation

**Objective**: the task addresses the problem of image search **result diversification** in the context of **social photo retrieval**.

Why diversifying search results?
- a method of tackling queries with unclear information needs;
- queries involve many declinations, e.g., sub-topics;
- widens the pool of possible results and increases the system performance;
- ...

**Relevance and Diversity** (~antinomic):
too much *diversification* may result in losing relevant items while increasing only the *relevance* will tend to provide near duplicate information.

---

Diversity Task: Objective & Motivation #2

The concept appeared initially for text retrieval but regains its popularity in the context of multimedia retrieval.

[Google Image Search for “Eiffel tower”, 12-10-2014]
Diversity Task: Use Case

To disambiguate the diversification need, we introduced a very focused use case scenario …

**Use case:** we consider a **tourist use case** where a person tries to find more information about a place she is potentially visiting. The person has only a vague idea about the location, knowing the name of the place.

… e.g., looking for **Rialto Bridge** in Italy

---

Diversity Task: Use Case #2

… learn more information from Wikipedia
Diversity Task: Use Case #3

… how to get some more accurate photos?

query using text “Rialto Bridge” …

… browse the results

Diversity Task: Use Case #4

![Image of Rialto Bridge photos from various sources]
Diversity Task: Use Case #5

... too many results to process,
inaccurate, e.g., people in focus, other views or places
meaningless objects
redundant results, e.g., duplicates, similar views ...
Diversity Task: Definition

Participants receive a ranked list of photos with locations retrieved from Flickr using its default “relevance” algorithm.

**Goal of the task:** refine the results by providing a ranked list of up to 50 photos (summary) that are considered to be both relevant and diverse representations of the query.

**relevant**: common photo representation of the location, e.g., different views at different times of the day/year and under different weather conditions, inside views, close-ups, drawings, sketches, creative views, which contain partially or entirely the target location.

**diverse**: depicting different visual characteristics of the location, with a certain degree of complementarity, i.e., most of the perceived visual information is different from one photo to another.

*we thank the task survey respondents for their precious feedback on these definitions.

Diversity Task: Target

going from this …
Diversity Task: Target

... to something like this:

Dataset: General Information

The dataset consists of 300 landmark locations (natural or man-made, e.g., sites, museums, monuments, buildings, roads, bridges) unevenly spread over 35 countries around the world:
Dataset: Resources

Location information consists of:
- the location name & GPS coordinates;
- a link to its Wikipedia web page;
- up to 5 representative photos from Wikipedia;
- a ranked set of Creative Commons photos retrieved from Flickr (up to 300 photos per location);
- metadata from Flickr (e.g., tags, description, views, #comments, date-time photo was taken, username, userid, etc);
- some general purpose visual and text content descriptors;
- an automatic prediction of user annotation credibility;
- relevance and diversity ground truth (up to 25 classes).

Retrieval method (we use Flickr API):
- use of the location name as query.

[2014: more focus on social aspects]

* the differences compared to 2013 data are depicted in bold.

Dataset: User Credibility

Idea: give an automatic estimation of the quality of tag-image content relationships;
~ indication about which users are most likely to share relevant images in Flickr (according to the underlying task scenario).

- visualScore: for each Flickr tag which is identical to an ImageNet concept, a classification score is predicted and the visualScore of a user is obtained by averaging individual tag scores;
- faceProportion: the percentage of images with faces out of the total of images tested for each user;
- uploadFrequency: average time between two consecutive uploads in Flickr;
...
Dataset: Statistics

Some basic statistics:

- **devset** (intended for designing and validating the methods)
  
<table>
<thead>
<tr>
<th>#locations</th>
<th>#images</th>
<th>min-average-max img. per location</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>8,923</td>
<td>285 - 297 - 300</td>
</tr>
</tbody>
</table>

- **testset** (intended for final benchmark)
  
<table>
<thead>
<tr>
<th>#locations</th>
<th>#images</th>
<th>min-average-max img. per location</th>
</tr>
</thead>
<tbody>
<tr>
<td>123</td>
<td>36,452</td>
<td>277 - 296 - 300</td>
</tr>
</tbody>
</table>

⇒ total number of provided images: 45,375.

- **credibilityset** (intended for training/designing credibility desc.)
  
<table>
<thead>
<tr>
<th>#locations</th>
<th>#images*</th>
<th>#users</th>
<th>average img. per user</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>3,651,303</td>
<td>685</td>
<td>5,330</td>
</tr>
</tbody>
</table>

* images are provided via Flickr URLs.

Dataset: Ground Truth

Relevance and diversity annotations were carried out by **expert annotators**:

- **devset**: relevance (3 annotations), diversity (1 annotation issued from 2 experts + 1 final master revision);

- **testset**: relevance (3 annotations issued from 11 expert annotators), diversity (1 annotation from 3 expert annotators + 1 final master revision);

- **credibilityset**: only relevance for 50,157 photos (3 annotations issued from 9 experts);

- lenient majority voting for relevance.

* advanced knowledge of location characteristics mainly learned from Internet sources.
Dataset: Ground Truth #2

Some basic statistics:

- **devset:**
  - relevance | Kappa agreement* | % relevant img.
  - diversity | avg. clusters per location | avg. img. per cluster
  - 0.85 | 23 | 8.9

- **testset:**
  - relevance | Kappa agreement* | % relevant img.
  - diversity | avg. clusters per location | avg. img. per cluster
  - 0.75 | 23 | 8.8

- **credibilityset:**
  - relevance | Kappa agreement* | % relevant img.
  - diversity
  - 0.75 | 23 | 8.8

*Kappa values > 0.6 are considered adequate and > 0.8 are considered almost perfect.

Dataset: Ground Truth #3

Diversity annotation example (Aachen Cathedral, Germany):

chandelier | architectural details | stained glass | windows | archway | mosaic | creative views | close up | mosaic | outside winter view


Evaluation: Required Runs

Participants are required to submit up to 5 runs:

- **required runs:**
  - run 1: automated using *visual information only*;
  - run 2: automated using *textual information only*;
  - run 3: automated using *textual-visual* fused without other resources than provided by the organizers;

- **general runs:**
  - run 4: automated using *credibility information*;
  - run 5: *everything allowed*, e.g., human-based or hybrid human-machine approaches, including using data from external sources (e.g., Internet).

Evaluation: Official Metrics

- **Cluster Recall** @ $X = \frac{Nc}{N}$ (CR@X)
  
  where $X$ is the cutoff point, $N$ is the total number of clusters for the current location (from ground truth, $N<=25$) and $Nc$ is the number of different clusters represented in the X ranked images;

- **Precision** @ $X = \frac{R}{X}$ (P@X)
  
  where $R$ is the number of relevant images;

- **F1-measure** @ $X = \text{harmonic mean of CR and P}$ (F1@X)

Metrics are reported for different values of $X$ (5, 10, 20, 30, 40 and 50) on per location basis as well as overall (average).

* cluster recall is computed only for the relevant images.
Participants: Basic Statistics

- **Survey (February 2014):**
  - 66 (55) respondents were interested in the task, 26 (23) very interested;

- **Registration (April 2014):**
  - 20 (24) teams registered from 15 (18) different countries (3 teams are organizer related);

- **Crossing the finish line (September 2014):**
  - 14 (11) teams finished the task, 12 (8) countries, including 3 organizer related teams (no late submissions);
  - 54 (38) runs were submitted from which 1 (2) brave human-machine!

- **Workshop participation (October 2013):**
  - 10 (8) teams are represented at the workshop.

* the numbers in the brackets are from 2013.

---

Participants: Submitted Runs

<table>
<thead>
<tr>
<th>team</th>
<th>country</th>
<th>1-visual</th>
<th>2-text</th>
<th>3-text-visual</th>
<th>4-cred.</th>
<th>5-free</th>
</tr>
</thead>
<tbody>
<tr>
<td>BU-Retina</td>
<td>Turkey</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>visual</td>
</tr>
<tr>
<td>CEALIST*</td>
<td>France, Austria</td>
<td>√</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>visual</td>
</tr>
<tr>
<td>DCLab</td>
<td>Hungary</td>
<td>√</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>DWS</td>
<td>Germany</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>LAPI*</td>
<td>Romania, Italy</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>miracl</td>
<td>Tunisia</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>TUW*</td>
<td>Austria</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>MIS</td>
<td>Austria</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>MMLab</td>
<td>Belgium, S. Korea</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>PeRceiVe@UNICT</td>
<td>Italy</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>PRe-MM</td>
<td>Italy</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Recod</td>
<td>Brazil</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>SocSens</td>
<td>Greece</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>UNED</td>
<td>Spain</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

* organizer related team.
Results: P vs. CR @20 (all runs)

- Flickr initial
- PRa-MM
- SocSens

<table>
<thead>
<tr>
<th>team/run</th>
<th>P@10</th>
<th>P@20</th>
<th>CR@10</th>
<th>CR@20</th>
<th>F1@10</th>
<th>F1@20</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRa-MM2014_run5_Final</td>
<td>0.8659</td>
<td>0.8512</td>
<td>0.2976</td>
<td>0.4692</td>
<td>0.4362</td>
<td>0.5971</td>
</tr>
<tr>
<td>SocSens2014_run5</td>
<td>0.8268</td>
<td>0.815</td>
<td>0.3027</td>
<td>0.4747</td>
<td>0.4394</td>
<td>0.5943</td>
</tr>
<tr>
<td>CEALIST2014_run_5_general</td>
<td>0.7951</td>
<td>0.7931</td>
<td>0.2803</td>
<td>0.4563</td>
<td>0.4076</td>
<td>0.571</td>
</tr>
<tr>
<td>TUW2014_RUN1_visual_new_test</td>
<td>0.7984</td>
<td>0.7887</td>
<td>0.2827</td>
<td>0.4497</td>
<td>0.4124</td>
<td>0.5602</td>
</tr>
<tr>
<td>LAPI2014_run2_HC-RF_text_TI</td>
<td>0.7984</td>
<td>0.7882</td>
<td>0.2861</td>
<td>0.4431</td>
<td>0.3628</td>
<td>0.5583</td>
</tr>
<tr>
<td>Run5_UNED2014all_results</td>
<td>0.7748</td>
<td>0.7772</td>
<td>0.2679</td>
<td>0.4343</td>
<td>0.3932</td>
<td>0.5502</td>
</tr>
<tr>
<td>MMLab2014_run5_useall</td>
<td>0.7967</td>
<td>0.8008</td>
<td>0.2508</td>
<td>0.4252</td>
<td>0.3748</td>
<td>0.5455</td>
</tr>
<tr>
<td>RECOD2014_dev+credibility</td>
<td>0.7439</td>
<td>0.7598</td>
<td>0.2585</td>
<td>0.4288</td>
<td>0.3805</td>
<td>0.5423</td>
</tr>
<tr>
<td>DCLab2014_run3_VisTextClusterAvgRelevance</td>
<td>0.7927</td>
<td>0.7756</td>
<td>0.2578</td>
<td>0.4127</td>
<td>0.3838</td>
<td>0.5305</td>
</tr>
<tr>
<td>PeRCeVe@UNICT2014_run2</td>
<td>0.7463</td>
<td>0.7553</td>
<td>0.2271</td>
<td>0.3902</td>
<td>0.3431</td>
<td>0.5063</td>
</tr>
<tr>
<td>BU-Image_visSIFT_R5</td>
<td>0.7203</td>
<td>0.7228</td>
<td>0.2339</td>
<td>0.387</td>
<td>0.3492</td>
<td>0.4966</td>
</tr>
<tr>
<td>MIS2014_run3</td>
<td>0.6748</td>
<td>0.6732</td>
<td>0.2336</td>
<td>0.3985</td>
<td>0.3433</td>
<td>0.4949</td>
</tr>
<tr>
<td>Flickr initial results</td>
<td>0.8086</td>
<td>0.8065</td>
<td>0.2112</td>
<td>0.3427</td>
<td>0.3287</td>
<td>0.4699</td>
</tr>
<tr>
<td>DWS2014_run1_visualTestset(2)</td>
<td>0.7715</td>
<td>0.7524</td>
<td>0.2224</td>
<td>0.3405</td>
<td>0.3385</td>
<td>0.46</td>
</tr>
<tr>
<td>MIRACL2014_run1_visualInformationOnly</td>
<td>0.8033</td>
<td>0.7772</td>
<td>0.2145</td>
<td>0.3285</td>
<td>0.3329</td>
<td>0.4501</td>
</tr>
</tbody>
</table>

Best improvements compared to Flickr (in percentage points): P@20 4.5, CR@20 13.
Results: Best Team Runs

*ranking based on official metrics (F1@20).

Results: Best Team Runs #2

*ranking based on official metrics (F1@20).
Results: Visual Results - Flickr Initial

Salisbury Cathedral
F1@20=0.3333,
P@20=1, CR@20=0.2, 25 clusters

Results: Visual Results #2 - Best F1@20

Salisbury Cathedral
F1@20=0.6721,
P@20=0.95, CR@20=0.52, 25 clusters
Results: Visual Results #3 - Lowest F1@20

Salisbury Cathedral
F1@20 = 0.3333, P@20=1, CR@20=0.2, 25 clusters

Brief Discussion

Methods:
- this year mainly clustering, re-ranking, optimization-based and relevance feedback (including machine-human);
- best run F1@20: pre-filtering + hierarchical clustering + tree refining + re-ranking using visual-text-cred. information (PRa-MM);
- user tagging credibility information proved its potential and should be further investigated in social retrieval scenarios.

Dataset:
- still low resources for location Creative Commons on Flickr;
- diversity annotation for 300 photos much difficult than for 100;
- descriptors were very well received (employed by most of the participants).
Present & Perspectives

For 2014:
- the task was a full task this year,
- the entire dataset is to be publicly released (soon).

For 2015:
- working on a new use case scenario.

Acknowledgements. Many thanks to:
MUCKE project (Mihai Lupu, Adrian Popescu) for funding the annotation process.
Task auxiliaries: Alexandru Gînscă (CEA LIST, France), Adrian Iftene (Faculty of Computer Science, Alexandru Ioan Cuza University, Romania).

Questions & Answers

Thank you!

... and please contribute to the task by uploading free Creative Commons photos on social networks! 😊

See you at the poster session and for the technical retreat ...