DISA at ImageCLEF 2014: The Search-based Solution for Scalable Image Annotation

Petra Budikova, Jan Botorek,
Michal Batko, Pavel Zezula

Outline

- DISA lab introduction
- Our search-based solution to annotation task
- New features
- Experimental results with the new features
- Conclusions

DISA Lab

- Laboratory of Data Intensive Systems and Applications
  - Leader: prof. Pavel Zezula
  - Masaryk university, Czech Republic
  - http://disa.fi.muni.cz/

- Focus on techniques for effective and efficient data management
  - Similarity-based indexing and searching
  - Content-based multimedia processing
  - Distributed data organization

ImageCLEF Scalable Image Annotation Task

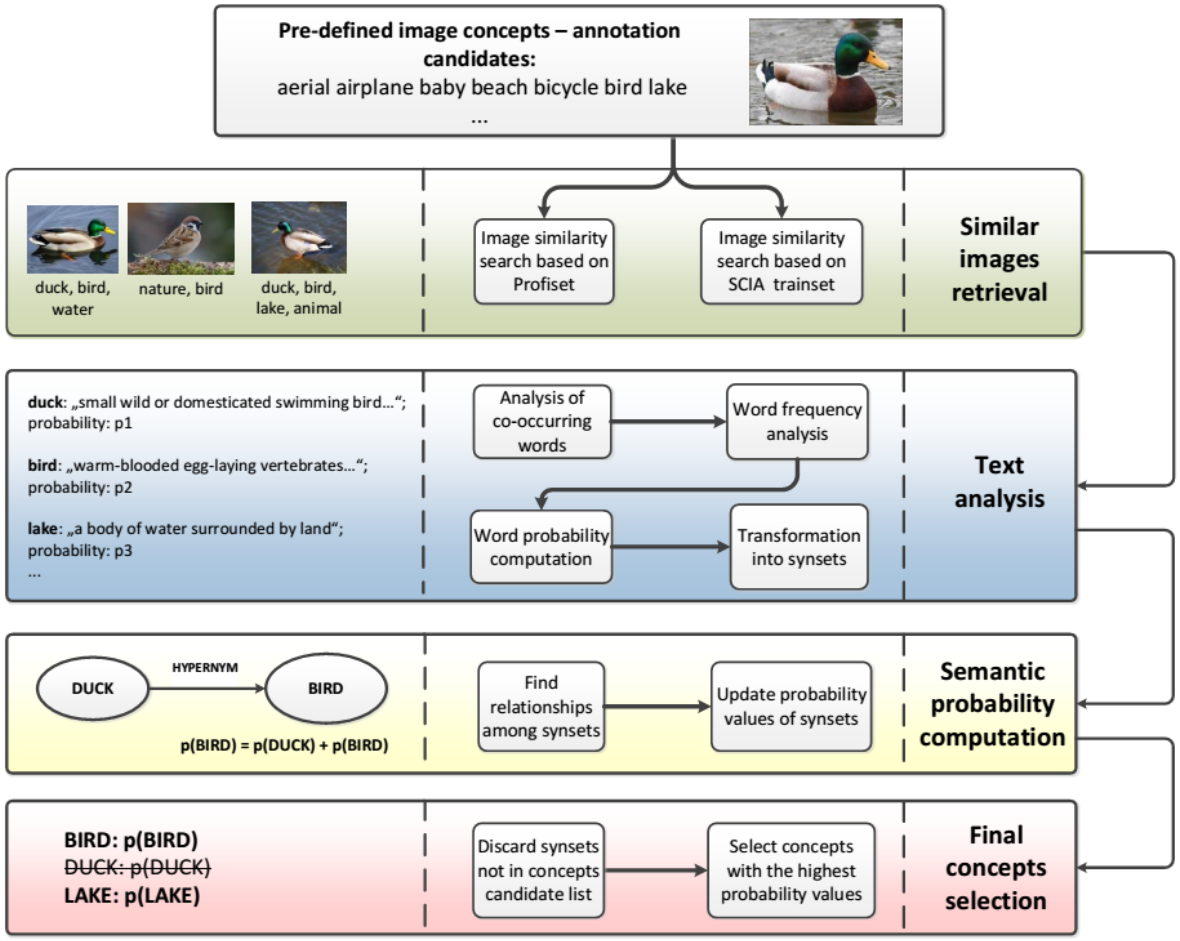
- Annotation task definition
  - Input: image + set of candidate concepts
  - Expected result: set of relevant concepts

aerial airplane baby beach bicycle bird boat bridge building car cartoon castle cat chair child church cityscape closeup cloud cloudless coast **countryside** **daytime** desert diagram dog drink drum elder embroidery fire firework fish flower fog food footwear furniture garden **grass** guitar harbor hat helicopter highway **horse** indoor instrument lake lightning logo monument moon motorcycle mountain nighttime overcast painting park person **plant** portrait protest rain rainbow reflection river road sand sculpture sea shadow sign silhouette smoke snow soil space spectacles sport sun sunrise/sunset table teenager toy traffic train tricycle truck underwater unpaved wagon water

- 2 datasets
  - Development data: 1940 images, ground truth available
  - Test data: 7291 images, ground truth not available
- Evaluation script provided by organizers – precision, recall, F-measure

Our solution

- Search-based annotation with utilization of semantic relationships defined by WordNet



Our solution (cont.)

- Image datasets for similarity-based searching:
  - Profiset: 20M images with high-quality keywords
  - Dataset provided by ImageCLEF organizers ("SCIA trainset"): 500K images from internet, descriptions more noisy, but covers all topics in the contest
- Image content extraction:
  - Combination of 5 MPEG7 global features
- Exploitation of semantic relationships:
  - Synonyms
  - Probability ranking of possible meanings of each word
  - Hypernymy/hyponymy
  - Holonymy/meronymy

New features for image retrieval

- $DeCAF_7$ visual features
  - Utilization of deep convolutional network
  - Outperformed all participants at ImageNet large scale visual recognition challenge ILSVRC-2012 (Krizhevsky et. al. 2012)
  - Adopted as visual descriptor (Donahue et. al. 2013)
    - Result from the last hidden layer used as 4096-dimensional visual descriptor
    - Similarity using classical L_p metric
    - Gives better results than traditional features on benchmarks from other domains

- Easily used by our similarity-search framework
  - PPP-Codes technique able to index 20M collection of data
  - Real-time response on a common server hardware
    - 8 cores, 8GB RAM, 256GB SSD

- Improved results of our annotation!

Evaluation results

- Development data

	mP-concept	mR-concept	mF-concept	mP-sample	mR-sample	mF-sample	mAP-sample
Baseline (random)	*0.0775*	*0.0641*	*0.0498*	*0.0730*	*0.0969*	*0.0722*	*0.1578*
DISA-best with MPEG and Profiset data	0.2954	0.2746	0.2184	0.3044	0.4516	0.3352	0.4268
DISA-best with MPEG and Profiset+SCIA data	**0.2919**	**0.2778**	**0.2202**	**0.3052**	**0.4533**	**0.3369**	**0.4281**
DISA-best with DeCAF and Profiset data	0.4768	0.4899	0.4165	0.4466	0.6152	0.4825	0.6105
DISA-best with DeCAF and Profiset+SCIA data	**0.4928**	**0.5085**	**0.4315**	**0.4534**	**0.6252**	**0.4901**	**0.6196**

- Test data

	mF-concept	mF-sample	mAP-sample
Baseline (random)	*0.026*	*0.035*	*0.088*
DISA-best with MPEG and Profiset data	0.154	0.279	0.316
DISA-best with MPEG and Profiset+SCIA data	0.191	0.297	0.343
Competition best	**0.547 (0.548)**	**0.377**	**0.368 (0.370)**
DISA-best with DeCAF and Profiset+SCIA data	**0.411**	**0.399**	**0.486**

Evaluated by ImageCLEF organizers as a favor after competition deadline

New result evaluation – details

	mF-concept	mF-sample	mAP-sample
DISA-MU 04 (DISA best in competition)	19.1 [17.5–21.8]	29.7 [29.2–30.3]	34.3 [33.8–35.0]
KDEVIR 09 (competition winner)	54.7 [50.9–58.3]	37.7 [37.0–38.5]	36.8 [36.1–37.5]
DISA-MU NEW	41.1 [38.3–44.2]	39.9 [39.3–40.5]	48.6 [47.9–49.3]

System	MAP-samples				MF-samples				MF-concepts				
	all	ani.	food	207	all	ani.	food	207	all	ani.	food	207	unseen
KDEVIR 9	36.8	33.1	67.1	28.9	37.70	29.9	64.9	32.0	54.7	67.1	65.1	31.6	66.1
DISA NEW	48.6	51.0	67.2	32.3	39.90	44.4	48.5	26.7	41.1	N/A	N/A	N/A	44.9
MIL 3	36.9	30.9	68.6	23.3	27.50	20.6	53.1	18.0	34.7	34.7	50.4	16.9	36.7
MindLab 1	37.0	43.1	63.0	22.1	25.80	17.0	45.2	18.3	30.7	35.1	35.3	16.7	34.7
MLIA 9	27.8	18.8	53.6	16.7	24.80	12.1	46.0	16.4	33.2	32.7	37.3	16.9	34.8
DISA 4	34.3	46.6	39.6	19.0	29.70	40.6	31.2	16.9	19.1	23.0	22.3	7.3	19.0
RUC 7	27.5	25.2	44.2	15.1	29.30	28.0	28.2	20.7	25.3	20.1	23.1	10.0	18.7
IPL 9	23.4	30.0	48.5	18.9	18.40	20.2	29.8	17.5	15.8	15.8	33.3	12.5	22.0
IMC 1	25.1	35.7	35.6	12.9	16.30	14.3	21.0	10.9	12.5	10.2	15.1	6.1	11.2
INAOE 5	9.6	6.9	15.0	8.5	5.30	0.4	0.5	6.4	10.3	1.0	0.8	17.9	19.0
NII 1	14.7	23.2	22.0	4.6	13.00	18.9	18.7	4.9	2.3	3.0	2.1	0.9	1.8
FINKI 1	6.9	N/A	N/A	N/A	7.20	8.1	12.3	4.1	4.7	6.3	9.0	2.9	4.7

Evaluation results – influence of semantic links

- Development data, similarity search on Profiset only

	mP-concept	mR-concept	mF-concept	mP-sample	mR-sample	mF-sample	mAP-sample
MPEG, basic frequency analysis	0.1824	0.3290	0.1904	0.2383	0.4083	0.2755	0.3467
MPEG, multiple meanings, no links	0.2912	0.2921	0.2240	0.2826	0.3953	0.3032	0.3838
MPEG, multiple meanings, hyper/hypo	0.2915	0.2667	0.2121	0.3008	0.4420	0.3306	0.4211
MPEG, multiple meanings, hyper/hypo and mero/holo	**0.2954**	**0.2746**	**0.2184**	**0.3044**	**0.4516**	**0.3352**	**0.4268**
Caffe, basic frequency analysis	0.3247	0.4684	0.3360	0.3735	0.4990	0.3962	0.4950
Caffe, multiple meanings, no links	0.4887	0.4881	0.4058	0.4268	0.5561	0.4488	0.5564
Caffe, multiple meanings, hyper/hypo	0.4803	0.4849	0.4149	0.4464	0.6096	0.4808	0.6076
Caffe, multiple meanings, hyper/hypo and mero/holo	**0.4768**	**0.4899**	**0.4165**	**0.4466**	**0.6152**	**0.4825**	**0.6105**

Conclusions

- Presented modular architecture of DISA annotation tool
  - allows easy replacement of any component

- Our approach is based on nearest-neighbor search not training
  - completely scalable – crawled data can be directly indexed
  - no need for ground truth
  - generic vocabulary (keyword) annotation – no need to hit predefined classes

- New visual similarity by DeCAF features
  - The new similarity-search component enabled us to increase the quality of annotations by approximately 10-20 % (depending on the quality measure)
  - New DISA results outperform the best results submitted to ImageCLEF 2014 Annotation Challenge in 2 out of 3 quality measures

Questions?

More information about the new feature results can be found here:

DISA at ImageCLEF 2014 Revised: Search-based Image Annotation with DeCAF Features.

Petra Budikova, Jan Botorek, Michal Batko, Pavel Zezula.

Technical Report. Computing Research Repository,

http://arxiv.org/abs/1409.4627