Learning Visual Entities and their Visual Attributes from Text Corpora

Erik Boiy
Koen Deschacht
Marie-Francine Moens

Department of Computer Science
Katholieke Universiteit Leuven
Belgium

Introduction

- Goal: Determine the visualness of words accompanying an image
  - Entities (nouns) and their attributes (adjectives)
- Methods:
  - Construct a dictionary of visual and non-visual words
  - Compared to knowledge-rich approach using WordNet
- Potential uses:
  - Automatically assign index descriptors to images
  - Automatically annotate images
  - Aligning different media

Overview

- Introduction
- ➔ Methodology
  - ➔ Corpus-based association techniques
  - WordNet-based approach
- Experiments, results and discussion
- Conclusions & Future work

Corpus-based association techniques

- Hypothesis testing
  - Which words are related to a specific domain?
  - States *independence* between a term and the domain
- Domain represented by a *target corpus*
  - Assumed to be visual
- Compared to a general *reference corpus*
  - Assumed to be non-visual

Likelihood ratio (1)

- Contingency table

	visual	!visual	
term = t	c_{12}	$c_2 - c_{12}$	c_2
term != t	$c_1 - c_{12}$	$N + c_{12} - c_1 - c_2$	$N - c_2$
	c_1	$N - c_1$	N

- $p_1 = P(term = t \mid visual)$ $p_2 = P(term = t \mid !visual)$
  - Assume a binomial distribution
- Define

$$p = \frac{c_2}{N} \qquad p_1 = \frac{c_{12}}{c_1} \qquad p_2 = \frac{(c_2 - c_{12})}{(N - c_1)}$$

- $H_1 : p_1 = p_2 = p$

Likelihood ratio (2)

- $L(H_1) = b(c_{12}; c_1, p)\, b(c_2 - c_{12}; N - c_1, p)$
  $L(H_2) = b(c_{12}; c_1, p_1)\, b(c_2 - c_{12}; N - c_1, p_2)$
- Define likelihood ratio λ

$$\lambda = \frac{L(H_1)}{L(H_2)} = \frac{L(p, c_{12}, c_1)\, L(p, c_2 - c_{12}, N - c_1)}{L(p_{1,} c_{12}, c_1)\, L(p_2, c_2 - c_{12}, N - c_1)}$$

- Where

$2[\log L(p1, c12, c1) + \log L(p2, c2 - c12, N - c1) - \log L(p, c12, c1) - \log L(p, c2 - c12, N - c1)]$¿

$$L(p, k, n) = p_k (1 - p)^{(n-k)}$$

- We take

$$\begin{aligned} -2\log\lambda = 2[&\log L(p_{1,} c_{12,} c_1) + \log L(p_{2,} c_2 - c_{12,} N - c_1) \\ &- \log L(p, c_{12,} c_1) - \log L(p, c_2 - c_{12,} N - c_1)] \end{aligned}$$

Pearson's chi-square test

- The χ² statistic is defined as

$$\chi^2 = \sum_{i,j} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}$$

with $E_{i,j}$ the expected value for $O_{i,j}$

- $E_{i,j}$ are calculated from the marginal probabilities

$$E_{i,j} = \frac{O_{i,1} + O_{i,2}}{N} \times \frac{O_{1,j} + O_{2,j}}{N} \times N$$

Overview

- Introduction
- ➔ Methodology
  - Corpus-based association techniques
  - ➔ WordNet-based approach
- Experiments, results and discussion
- Conclusions & Future work

WordNet

- A large lexical database of English
- Contains nouns, verbs, adjectives and adverbs
- Grouped into sets of cognitive synonyms (*synsets*), each expressing a distinct concept
- Synsets are interlinked by means of conceptual-semantic and lexical relations.
  - **nouns:** *hypernym/hyponym* (organizes synsets in a hierarchical tree)
  - **adjectives/nouns:** *attribute*
- Short definition of synset is provided

WordNet-based approach (1)

- Visualness (*vis*) := degree that an adj. or noun is considered visual
- Manually identify seed synsets (s_i)
  - "person", "red" (vis = 1)
  - "power", "confidential" (vis = 0)
- A synset close to a visual synset gets high visualness and vice versa

$$vis(s) = \sum_i vis(s_i) \frac{sim(s, s_i)}{C(s)}$$

$$C(s) = \sum_i sim(s, s_i)$$

WordNet-based approach (2)

- Noun similarity:

$$sim(S_1, S_2) = \frac{2 \log P(S_p)}{\log P(S_1) + \log P(S_2)}$$

- With: S_p the most specific synset that is a parent of S_1 and S_2
  $P(S_i)$ the probability of labeling any word in a text with (a descendant from) synset S_i

- Adjective similarity:
  - Compare overlap between definitions (*glosses*)
  - Expanded by the *attribute* relation

Combined approach

- Use the dictionary built by the corpus-based association techniques to guide the selection of the seed synsets

Overview

- Introduction
- Methodology
  - Corpus-based association techniques
  - WordNet-based approach
- ➔ Corpora
- Experiments, results and discussion
- Conclusions & Future work

Corpora: training

- Flower corpus
  - Flower descriptions
  - 15,226 word tokens
- Antiques corpus
  - Old picture collection
  - Source: Oregon archives
  - 619,515 word tokens
- English Wikipedia corpus
  - All articles: 407,074,407 word tokens
  - Subset (major religions in China): 16,965 word tokens

“African violets (Saintpaulia ionantha) are small, flowering houseplants or greenhouse plants belonging to the Gesneriaceae family. They are perhaps the most popular and most widely grown houseplant. Their thick, fuzzy leaves and abundant blooms in soft tones of violet, purple, pink, and white make them very attractive. Numerous varieties and hybrids are available. African violets grow best in indirect sunlight.”

“A small girl looks up at a person dressed in the costume of an animal which could be “Woody Woodchuck” at the State Fair in Salem, Oregon.”

Ground truth

- Art corpus
  - Source: Dayton Art Institute
  - Describes a collection of art items
  - 8 art items annotated:
    headnouns and their adjectives
  - 1,337 word tokens
- Classification
  - Dictionary lookup
  - WordNet-based: apply cutoff
- Evaluation
  - Accuracy, precision, recall, F_1-measure

The Yoruba are one of sub-Saharan Africa's oldest surviving cultures, with origins that can be traced back about a thousand years. Located pre-dominately in Nigeria, the Yoruba are known for their diverse and creative artistic production. [...]

These small sculptures depict two identical human figures. The wooden bodies are weathered brown and the hair is faded blue. Both sculptures have a round base about one inch high. [...]

Overview

- Introduction
- Methodology
  - Corpus-based association techniques
  - WordNet-based approach
- Corpora
- ➔ Experiments, results and discussion
- Conclusions & Future work

Results: corpus-based

- Nouns

Training corpora		α	? %	A %	Precision %		Recall %		F_1 %	
V	!V				V	!V	V	!V	V	!V
flowers	religion	-	52.55	**61.22**	81.71	**47.56**	**50.95**	79.59	**62.76**	**59.54**
flowers	religion	0.99	52.55	38.54	82.35	36.64	5.32	97.96	10.00	53.33
+ antiques	wiki	-	0.93	47.66	83.78	35.02	31.10	86.05	45.37	49.78
+ antiques	wiki	0.99	0.93	47.55	**84.02**	35.01	30.77	86.43	45.04	49.83

- Adjectives

Training corpora		α	? %	A %	Precision %		Recall %		F_1 %	
V	!V				V	!V	V	!V	V	!V
flowers	religion	-	24.69	**77.13**	**88.26**	**53.45**	**80.15**	68.13	**84.01**	**59.90**
flowers	religion	0.99	24.69	38.84	87.88	27.95	21.32	**91.21**	34.32	42.78
+ antiques	wiki	-	1.66	57.38	85.79	38.38	48.22	80.15	61.74	51.90
+ antiques	wiki	0.99	1.66	57.17	85.71	38.25	47.93	80.15	61.48	51.78

- Adjectives 1.7 times more likely to be visual than nouns
- 78% of adjectives also present in visual corpus vs. 57% of nouns

Results: WordNet-based

- Manual seed synset selection (noun/adj.)

Cutoff	? %	A %	Precision %		Recall %		F_1 %	
			V	!V	V	!V	V	!V
.3	0.00	64.00	82.02	43.87	62.02	68.58	70.63	53.51
.3	0.00	56.22	81.13	36.67	50.15	71.22	61.98	48.41

- Combined approach (noun/noun/adj./adj.)

Training corpora		Cutoff	? %	A %	Precision %		Recall %		F_1 %	
V	!V				V	!V	V	!V	V	!V
flowers	religion	.5	52.55	60.24	78.74	**46.61**	52.09	**74.83**	62.70	**57.44**
+ antiques	wiki	.3	0.93	**66.71**	**82.67**	46.42	**66.22**	67.83	**73.54**	55.12
flowers	religion	.3	24.69	**73.00**	80.00	**45.21**	**85.29**	36.26	**82.56**	40.24
+ antiques	wiki	.3	1.66	61.39	**87.80**	41.26	53.25	**81.62**	66.30	**54.81**

Results: Inheritance

- Chunking identifies noun phrases
- Within a chunk
  - A noun inherits the visualness of its modifying adj. (? 65.39%)
  - An adj. inherits the visualness of the noun it modifies (? 9.05%)

A %	Precision %		Recall %		F_1 %	
	V	!V	V	!V	V	!V
56.79	81.82	32.97	53.73	65.59	64.86	43.88
58.86	90.23	33.73	52.17	81.16	66.12	47.66

Conclusions & Future work

- Hypothesis testing is valuable to determine the visualness of a term
  - When trained on well suited corpora
  - On its own (especially for adj.)
  - To improve a WordNet-based method
- Future: integrate our approaches in text-based image retrieval models