Words and pictures: basic methods

D.A. Forsyth, UIUC
with: Kobus Barnard, U.Arizona; Pinar Duygulu, Bilkent U.; Nando de Freitas, UBC; Tamara Berg, UIUC; Derek Hoiem, UIUC; Ian Endres, UIUC; Ali Farhadi, UIUC; Gang Wang, UIUC;

Core Problems and Algorithms

- Problems:
  - Auto-annotation
    - predict words from pictures
  - auto-illustration
    - predict pictures from words
  - layout
    - use word/picture information to produce useful browsable structures
- Methods
  - Implicit association between words and picture structures
  - Explicit association between words and picture structures

An Implicit Association method

- Idea:
  - produce a joint probability model that produces both regions and words
  - link implicitly by mixing over multiple local models
    - hierarchical
      - common regions linked to common words
      - *then*
        - uncommon regions linked to uncommon words

Input

"This is a picture of the sun setting over the sea with waves in the foreground"

Image processing*

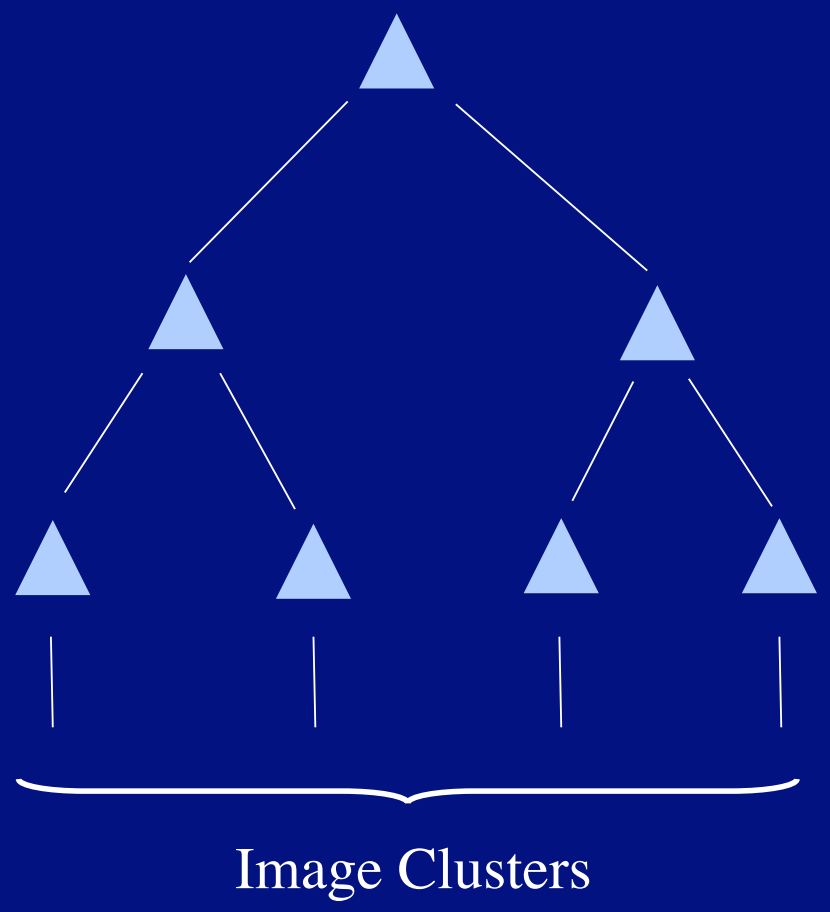
Each blob is a large vector of features

- Region size
- Position
- Colour
- Oriented energy (12 filters)
- Simple shape features

Language processing

sun sky waves sea

* Thanks to Blobworld team [Carson, Belongie, Greenspan, Malik], N-cuts team [Shi, Tal, Malik]



[Hofmann 98; Hofmann & Puzicha 98]

Node Behavior

Each node ....

Emits each modeled word, W , with some probability

Generates blobs according to a Gaussian distribution (parameters differ for each node).

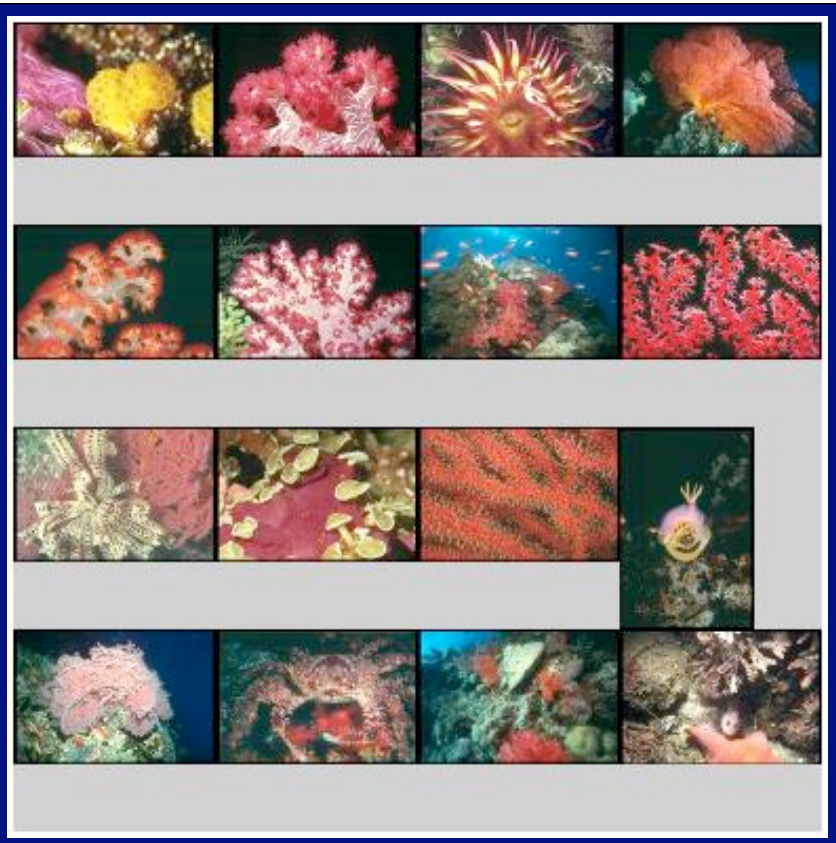
Nodes closer to the root emit more general / common words/blobs

Clustering algorithm

- Straightforward missing data problem
  - Missing data is path, nodes that generated each data element
- EM
  - If path, node were known for each data element, easy to get maximum likelihood estimate of parameters
  - given parameter estimate, path, node easy to figure out

Cluster found using only text

Cluster found using only blob features



Clusters found using both text and blob features

FAMSF Data

Web number: 4359202410830012

rec number: 2

Title: Le Matin

Primary class: Print

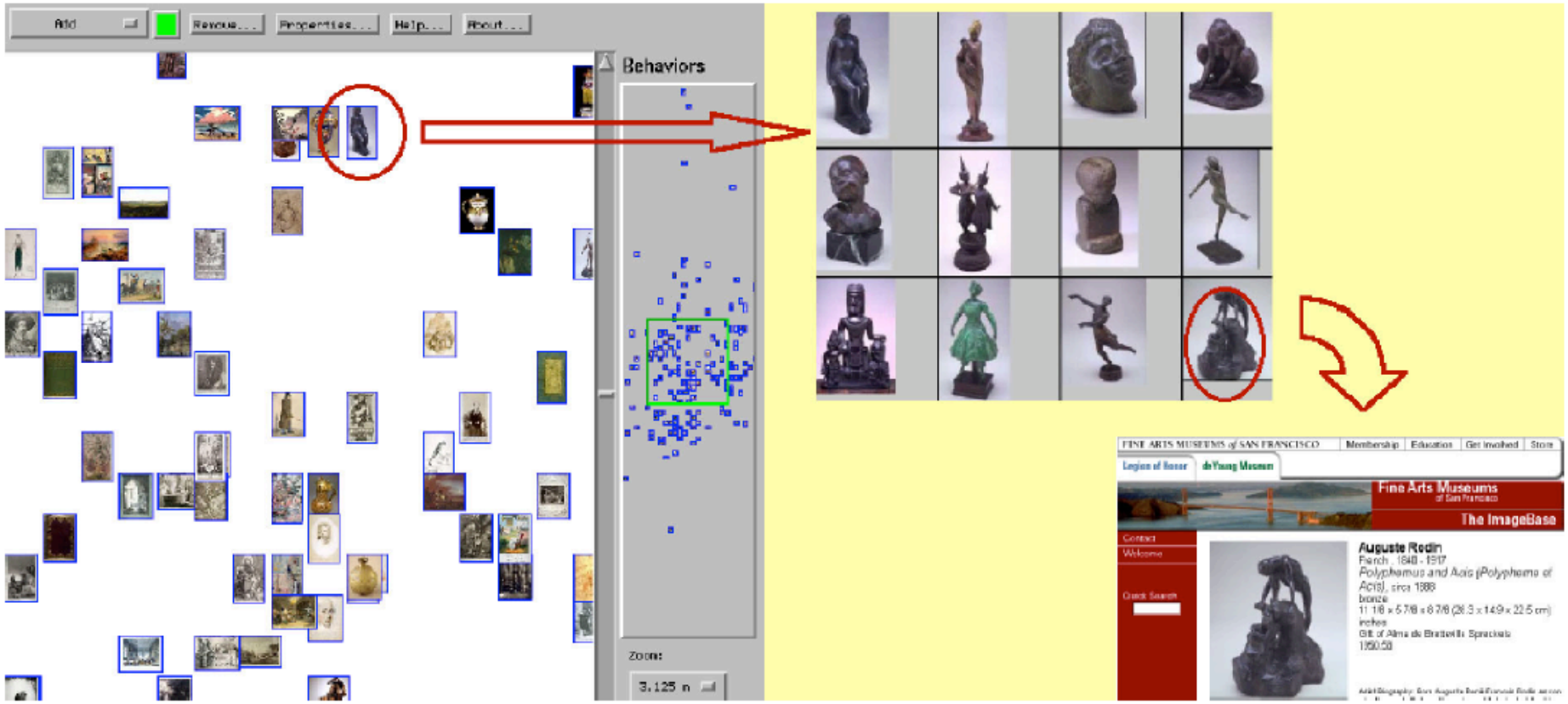
Artist: Tissot

Description:
serving woman stands in a dressing room, in front of vanity with chair, mirror and mantle, holding a tray with tea and toast

Display date: 1886

Country: France

83,000 images online, we clustered 8000


Add
Revoue...
Properties...
Help...
About...
Behaviors
Zoom:
3.125 n
FINE ARTS MUSEUMS of SAN FRANCISCO
Membership
Education
Get Involved
Store
Legion of Honor
deYoung Museum
Fine Arts Museums
of San Francisco
The ImageBase
Contact
Welcome
Quick Search
Auguste Rodin
French, 1840 - 1917
Polyphemus and Acis (Polyphème et Acis), circa 1889
bronze
11 1/8 x 5 7/8 x 8 7/8 (28.3 x 14.9 x 22.5 cm)
inches
Gift of Alma de Bretteville Spreckels
1950.58

Pictures from Words (Auto-illustration)

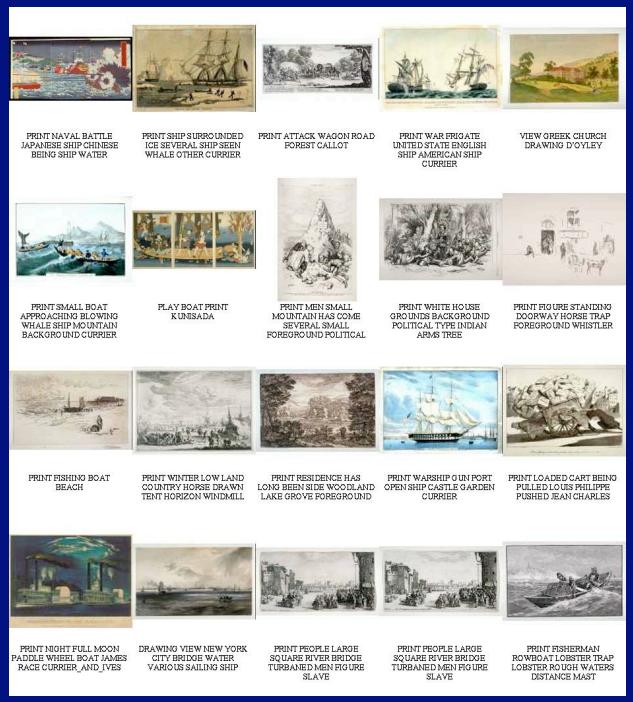
Text Passage (Moby Dick)

"The large importance attached to the harpooneer's vocation is evinced by the fact, that originally in the old Dutch Fishery, two centuries and more ago, the command of a whale-ship …"

Extracted Query

large importance attached fact old dutch century more command whale ship was person was divided officer word means fat cutter time made days was general vessel whale hunting concern british title old dutch ...

Retrieved Images



PRINT NAVAL BATTLE JAPANESE SHIP CHINESE BEING SHIP WATER

PRINT SHIP SURROUNDED ICE SEVERAL SHIP SEEN WHALE OTHER CURRIER

PRINT ATTACK WAGON ROAD FOREST CALLOT

PRINT WAR FRIGATE UNITED STATE ENGLISH SHIP AMERICAN SHIP CURRIER

PRINT SMALL BOAT APPROACHING BLOWING WHALE SHIP MOUNTAIN BACKGROUND CURRIER

PLAY BOAT PRINT KUNISADA

PRINT MEN SMALL MOUNTAIN HAS COME SEVERAL SMALL FOREGROUND POLITICAL

PRINT WHITE HOUSE GROUNDS BACKGROUND POLITICAL TYPE INDIAN ARMS TREE

Auto-annotation

- Predict words from pictures
  - Obstacle:
    - Hoffman's model uses document specific level probabilities
  - Dodge
    - smooth these empirically
- Attractions:
  - easy to score
  - large scale performance measures (how good is the segmenter?)
  - possibly simplify retrieval (Li+Wang, 03)


Keywords
GRASS TIGER CAT FOREST
Predicted Words (rank order)

tiger cat grass people water bengal buildings ocean forest reef

---


Keywords
HIPPO BULL mouth walk
Predicted Words (rank order)

water hippos rhino river grass reflection one-horned head plain sand

---


Keywords
FLOWER coralberry LEAVES PLANT

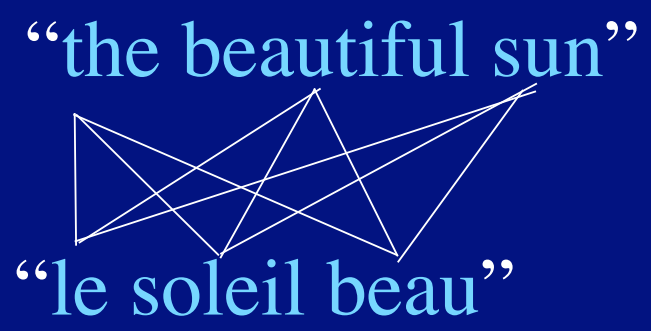
Predicted Words (rank order)
fish reef church wall people water landscape coral sand trees

An Explicit Association method

- Idea:
  - produce a joint probability for regions and words
    - vector quantize regions
    - if we knew which region produced which word, count

Machine Translation

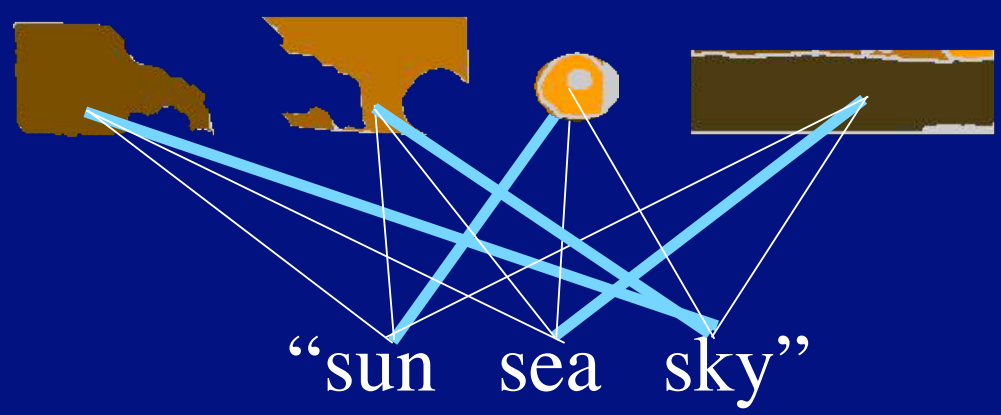
- Build a lexicon, produce MAP sentence in new language
- Lexicon building from an aligned bitext

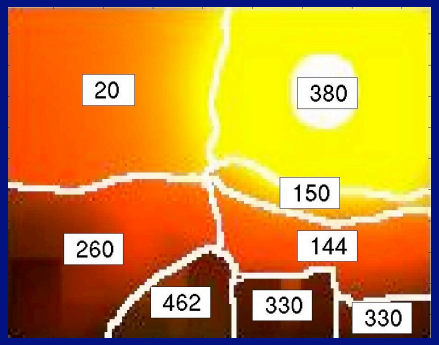


Brown, Della Pietra, Della Pietra & Mercer 93; Melamed 01

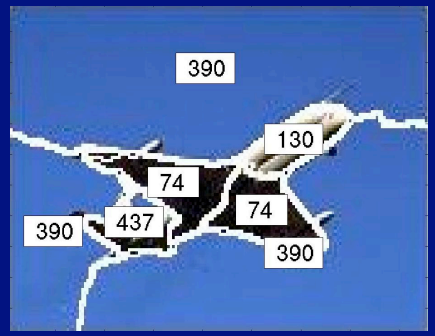
Lexicon building

- In its simplest form, missing variable problem
- Pile in with EM
  - given correspondences, conditional probability table is easy (count)
  - given cpt, expected correspondences could be easy
- Caveats
  - might take a lot of data; symmetries, biases in data create issues





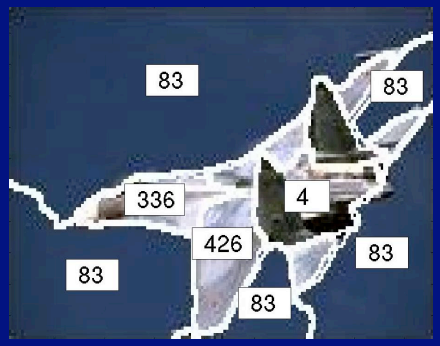
city mountain sky sun



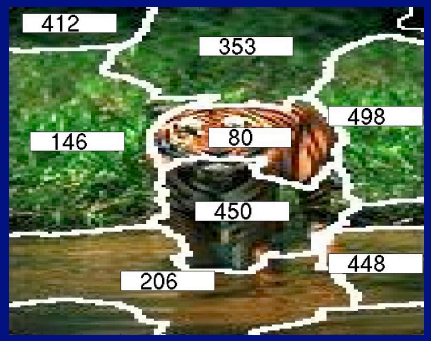
jet plane sky

cat forest grass tiger

beach people sun water



jet plane sky



cat grass tiger water

“Lexicon” of “meaning”

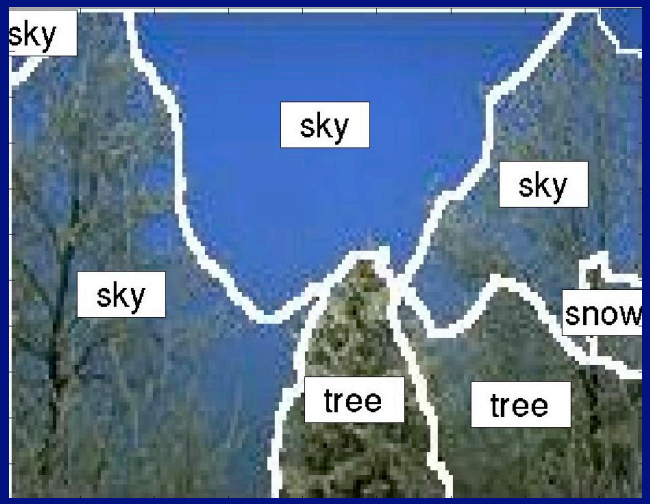
sun

sky

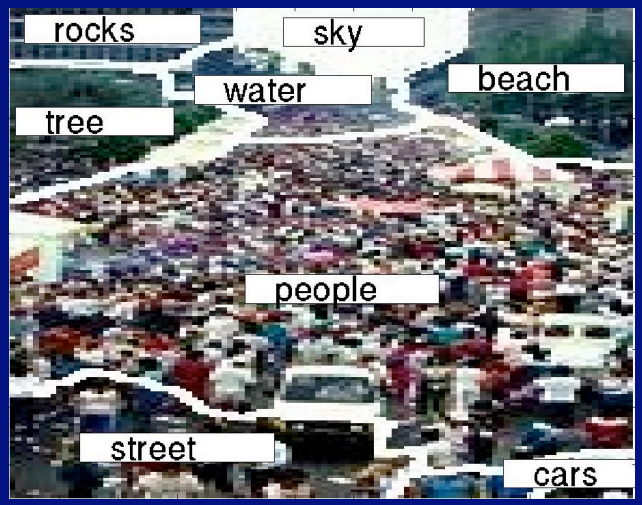
cat

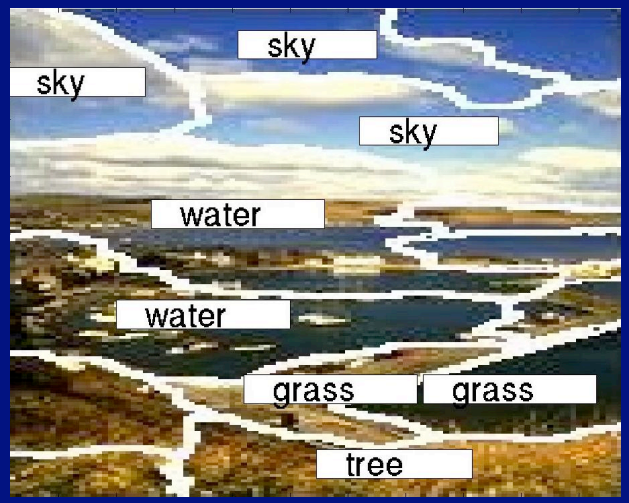
horse

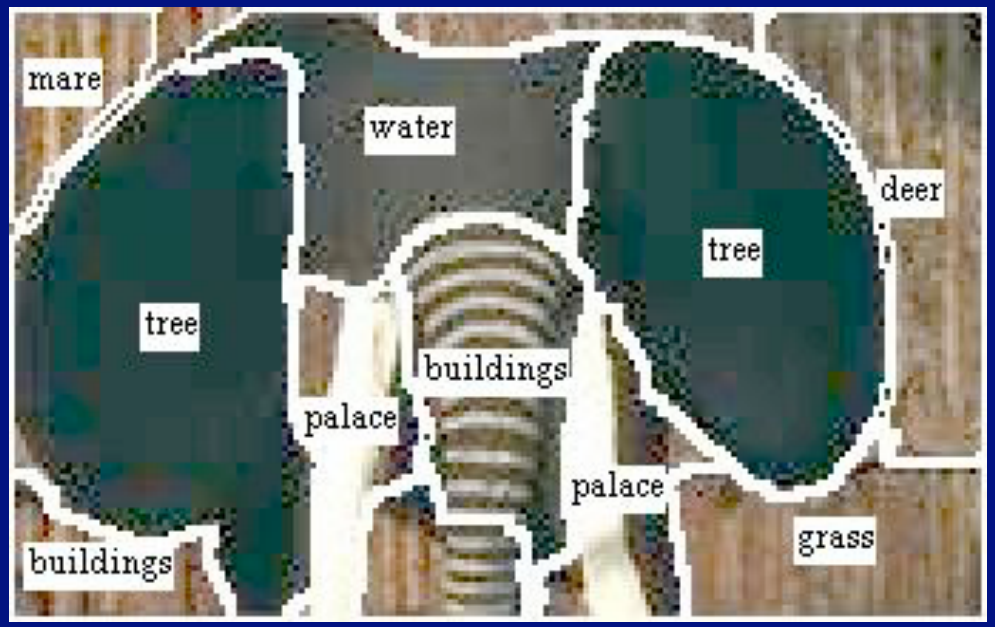
This could be either a conditional probability table or a joint probability table; each has significant attractions for different applications


sky
sky
sky
sky
snow
tree
tree


sky
sky
water
buildings
sky
tree
buildings
buildings


rocks
sky
water
beach
tree
people
street
cars


sky
sky
sky
water
water
grass
grass
tree

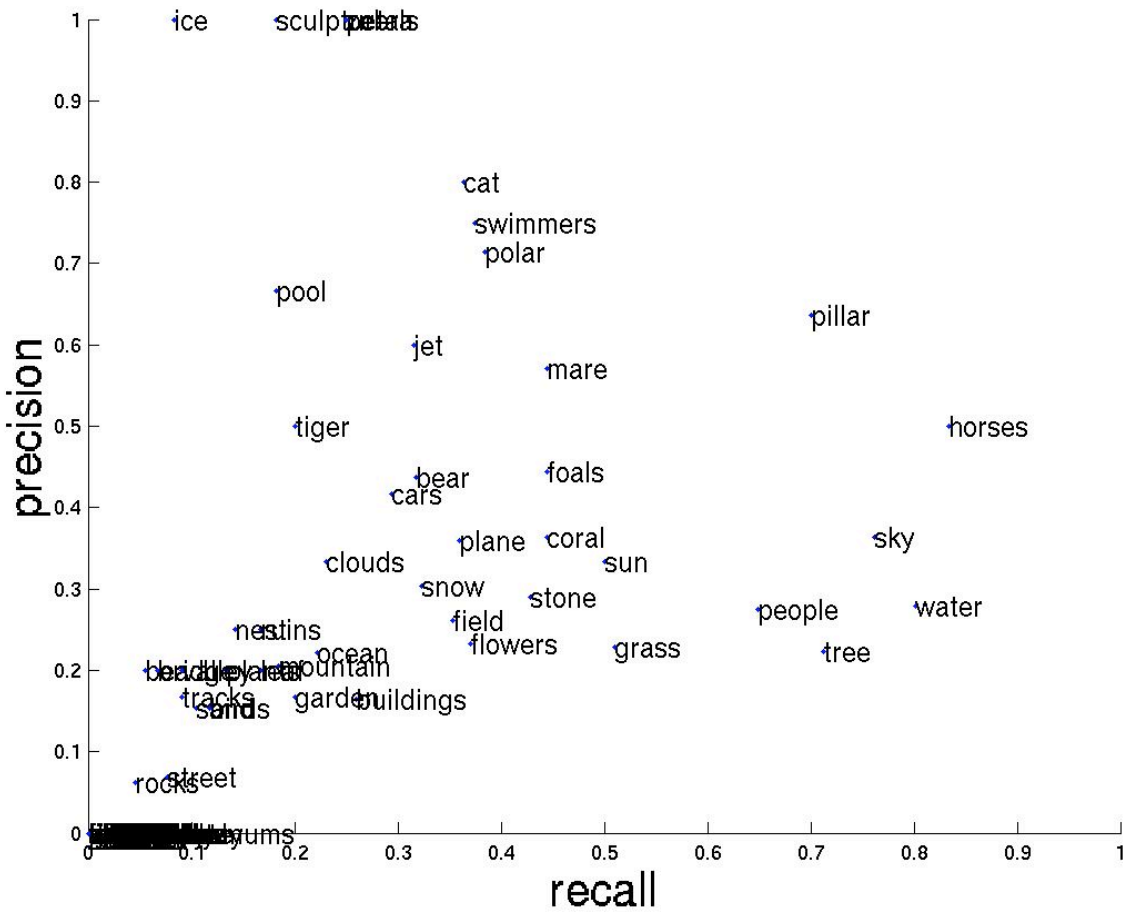

mare
water
deer
tree
tree
buildings
palace
palace
grass
buildings

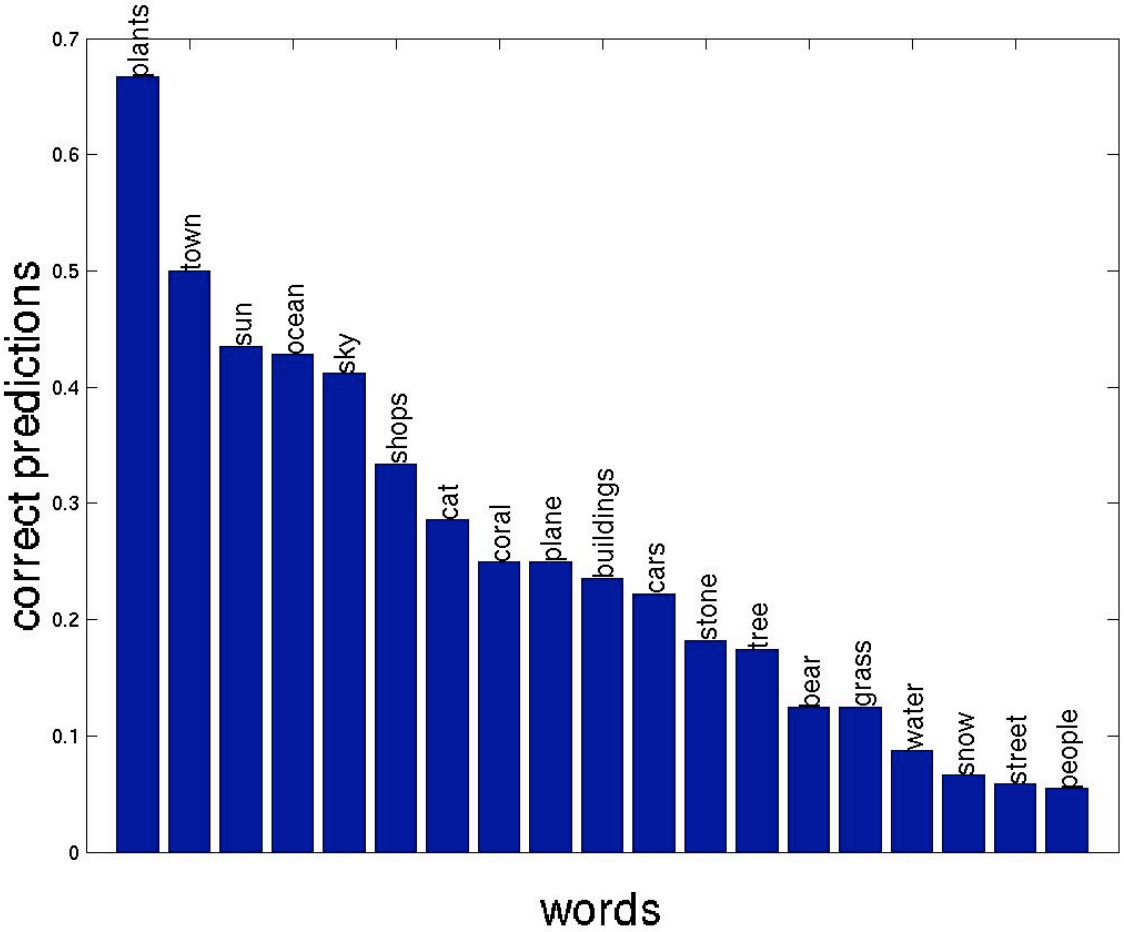
Performance measurement

By hand

By proxy

Grass Cat Buildings
Horses Tiger Mare


precision
recall
ice
sculpture
petals
cat
swimmers
polar
pool
pillar
jet
mare
tiger
horses
bear
foals
cars
plane
coral
sky
clouds
sun
snow
stone
people
water
nests
field
ocean
flowers
grass
tree
mountain
tracks
gardens
buildings
sands
rocks
street
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


correct predictions
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
plants
town
sun
ocean
sky
shops
cat
coral
plane
buildings
cars
stone
tree
bear
grass
water
snow
street
people
words

Datasets

- Matching words and pictures
  - http://kobus.ca/research/data/jmlr_2003/index.html
- Object recognition as machine translation (Corel-5K)
  - http://kobus.ca/research/data/eccv_2002/index.html

Accuracy and improvements

Method	P	R	F1	Ref	
Co-occ	0.03	0.02	0.02	[53]	Y. Mori et al 99
Trans	0.06	0.04	0.05	[27]	Duygulu et al, 02
CMRM	0.10	0.09	0.10	[37]	Jeon et al 03
TSIS	0.10	0.09	0.10	[19]	Celebi et al 05
MaxEnt	0.09	0.12	0.10	[39]	Jeon et al 04
CRM	0.16	0.19	0.17	[44]	Lavrenko et al 03
CT-3×3	0.18	0.21	0.19	[82]	Yavlinsky et al, 05
CRM-rect	0.22	0.23	0.23	[31]	Feng et al 04
InfNet	0.17	0.24	0.23	[50]	Metzler et al 04
MBRM	0.24	0.25	0.25	[31]	Feng et al 04
MixHier	0.23	0.29	0.26	[17]	Carneiro et al, 05
PicSOM	0.35*	0.35*	0.35*	[73]	Viitaniemi et al 07

More words

- Easy case
  - learn with larger vocabularies
  - tricky bits, but...
- Hard case
  - what do we do about out-of-example words?
  - one simple answer doesn’t work (later)

Example, pictures from Dan Kersten