

Manual resources like WordNet are noisy with limited coverage

- Cluster Nouns using Distributional Similarity
- Better Object Recognition

Manual resources like WordNet are noisy with limited coverage

- Cluster Nouns using Distributional Similarity
- Better Object Recognition
- Cluster ADJ modifiers (which modify objects)
  - red, green, wooden, rusty and furry
- Better Attribute Classifiers

Manual resources like WordNet are noisy with limited coverage

- Cluster Nouns using Distributional Similarity
- Better Object Recognition
  
- Cluster ADJ modifiers (which modify objects)
  - red, green, wooden, rusty and furry
- Better Attribute Classifiers
  
- Bootstrapping
  - Learning attribute values
    - **Color**: red, green, blue
    - **Size**: tall, big, huge

Manual resources like WordNet are noisy with limited coverage

- Cluster Nouns using Distributional Similarity
- Better Object Recognition
- Cluster ADJ modifiers (which modify objects)
  - red, green, wooden, rusty and furry
- Better Attribute Classifiers
- Bootstrapping
  - Learning attribute values
    - Color: red, green, blue
    - Size: tall, big, huge
  - Identify Visual and Non Visual nouns and adjectives
    - Visual: party
    - Non Visual: idea

## Raw Text Corpus (22 *billion* tokens)

- Gigaword corpus [Graff, 2003]
- Copy of news web crawled by [Ravichandran et al., 2005]

## Raw Text Corpus (22 billion tokens)

- Gigaword corpus [Graff, 2003]
- Copy of news web crawled by [Ravichandran et al., 2005]

## Part of speech Tagged Corpus (4 billion tokens)

- Web-derived ukWaC corpus  
`http://wacky.sslmit.unibo.it/doku.php`
- 2009 dump of the English Wikipedia  
`http://wacky.sslmit.unibo.it/doku.php`
- The New York Times section of the Gigaword
  - POS-tagged using TagChunk  
`http://www.umiacs.umd.edu/~hal/TagChunk/`

# Bag of Words as Context

7 words to left and right. Use raw text Gigaword + News Web corpus  
Contexts Weight: **Log Likelihood Ratio** Similarity: **Cosine**

animal

**shelter**  
shelters  
**cruelty**  
**veterinarians**  
experimentation  
animals  
**rescue**  
welfare  
rescues  
companion  
pets  
plant  
pet  
recipients  
humans

airplane

airplanes  
aircraft  
helicopter  
pilots  
**hangar**  
plane  
flight  
**hobby**  
planes  
jetliner  
**cockpit**  
thomas  
aviation  
airbus  
attendants

building

blocks  
**construction**  
buildings  
**renovation**  
facade  
**demolition**  
complexes  
**elevator**  
brick  
contractor  
concrete  
estimating  
permits  
downtown  
basement

# Collocation as Context

2 words to left and right with position. Use raw text Gigaword

Contexts Weight: [Pointwise Mutual Information](#) Similarity: [Cosine](#)

animal

bird  
animals  
dog  
elephant  
livestock  
fish  
wildlife  
child  
birds  
pig  
horse  
cattle  
patient  
whale  
species

airplane

plane  
aircraft  
airplanes  
jet  
airliner  
jetliner  
helicopter  
balloon  
car  
vehicle  
airline  
bus  
craft  
employer  
truck

building

buildings  
apartment  
rebuilding  
tower  
compound  
mall  
courthouse  
barracks  
premises  
structure  
warehouse  
houses  
cleaning  
complex  
roof



# ADJ modifiers as Context

Use JJ NN constructs. POS Tagged Data

Contexts Weight: [Pointwise Mutual Information](#) Similarity: [Cosine](#)

animal

animals  
mammals  
beasts  
mammal  
creatures  
pet  
beast  
creature  
prey  
bird  
pets  
sheep  
dogs  
birds  
insects

airplane

airplanes  
aeroplanes  
aeroplane  
helicopter  
monoplane  
aircraft  
airliner  
glider  
planes  
jet  
plane  
biplane  
craft  
fighter  
automobile

horse

horses  
stallion  
pony  
mare  
dog  
filly  
people  
racehorse  
time  
cow  
dogs  
ponies  
cat  
rider  
bull

# Nouns as Context for Adjective Clustering

Use JJ NN constructs. POS Tagged Data

wooden

Wooden  
rickety  
concrete  
wrought-iron  
movable  
sturdy  
glass-topped  
oaken  
collapsible  
moveable  
ornate  
cast-iron  
folding  
makeshift  
stainless

rusty

rusted  
stainless  
spring-loaded  
folding  
plated  
shiny  
wrought-iron  
galvanized  
heavy-duty  
removable  
wooden  
rusting  
yellowish  
sturdy  
large

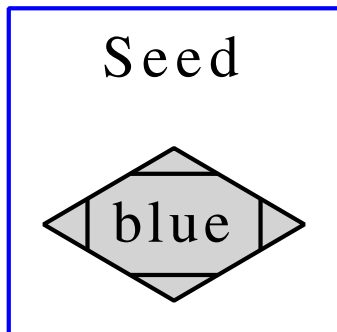
blue

red  
white  
black  
green  
maroon  
orange  
Blue  
beige  
yellow  
crimson  
sky-blue  
khaki  
striped  
blue-green  
coloured

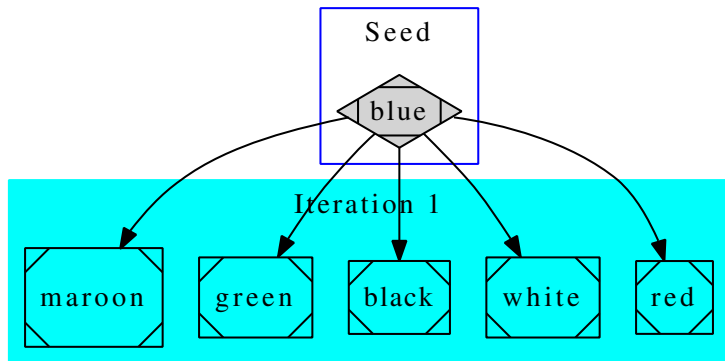
- Graph is constructed
  - Nodes: Adjectives
  - Neighbors: Top 50 similar adjectives
  - Edges: Distributional similarity as weights
- Seeds
  - Example Color: blue purple maroon beige green
- In-Degree (inDgr) [Kozareva et al., 2008]
- Mutual Exclusion [Thelen and Riloff, 2002]

# Semantic Classes of Attributes

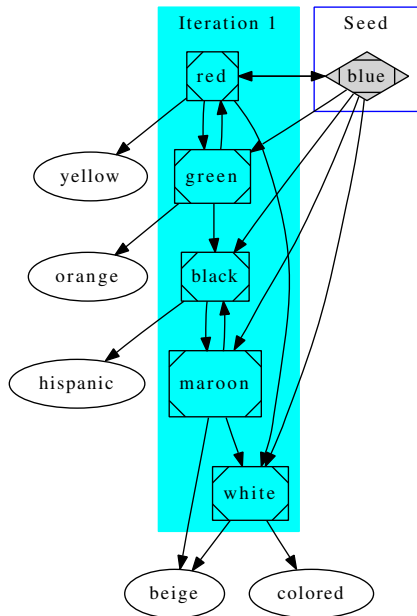
<b>Color:</b>	purple	blue	maroon	beige	green
<b>Material:</b>	plastic	cotton	wooden	metallic	silver
<b>Shape:</b>	circular	square	round	rectangular	triangular
<b>Size:</b>	small	big	tiny	tall	huge
<b>Surface:</b>	coarse	smooth	furry	fluffy	rough
<b>Direction:</b>	sideways	north	upward	left	down
<b>Pattern:</b>	striped	dotted	checked	plaid	quilted
<b>Quality:</b>	shiny	rusty	dirty	burned	glittery
<b>Beauty:</b>	beautiful	cute	pretty	gorgeous	lovely
<b>Age:</b>	young	mature	immature	older	senior
<b>Ethnicity:</b>	french	asian	american	greek	hispanic



# Bootstrapping in Action



# Bootstrapping in Action



- **Two free parameters in the bootstrapping algorithm**

- 1 Number of iterations (*iter*)
- 2 Number of new instances to add in each iteration (*m*)



# Evaluation of Semantic Classes of Attributes

## ■ Two free parameters in the bootstrapping algorithm

- 1 Number of iterations ( $iter$ )
- 2 Number of new instances to add in each iteration ( $m$ )

## ■ Build 10 systems ( $iter;m$ ) with different $iter$ and $m$

10;10	10;25	10;5	1;100
1;50	25;10	25;5	50;5
5; 25	5;50		

# Evaluation of Semantic Classes of Attributes

## ■ Two free parameters in the bootstrapping algorithm

- 1 Number of iterations (*iter*)
- 2 Number of new instances to add in each iteration (*m*)

## ■ Build 10 systems (*iter;m*) with different *iter* and *m*

10;10	10;25	10;5	1;100
1;50	25;10	25;5	50;5
5; 25	5;50		

## ■ Information Retrieval Style Human Evaluation

- Precision (Pr), Recall (Rec) and F-measure (F1)

# Human Evaluation of Attribute Classes

- Three Human Evaluators

- Evaluation GuideLines

- If something belongs to a semantic class

- For example: Color: red, green, white

- If something can be associated with a semantic class

- For example: Color: colorful, multicolor

# Human Evaluation of Attribute Classes

- Three Human Evaluators

- Evaluation GuideLines

- If something belongs to a semantic class

- For example: Color: red, green, white

- If something can be associated with a semantic class

- For example: Color: colorful, multicolor

- Inter Annotator Agreement ( $\kappa$ )

H1&H2	H2&H3	H1&H3
.45	.48	.48

Age is very subjective

# Human Evaluation of Attribute Classes

- Three Human Evaluators

- Evaluation GuideLines

- If something belongs to a semantic class

- For example: Color: red, green, white

- If something can be associated with a semantic class

- For example: Color: colorful, multicolor

- Inter Annotator Agreement ( $\kappa$ )

H1&H2	H2&H3	H1&H3
.45	.48	.48

H1&H2	H2&H3	H1&H3
.55	.59	.57

Age is very subjective

# Information Retrieval Style Evaluation

Evaluated 10 systems ( $iter;m$ ) with different  $iter$  and  $m$

10;10	10;25	10;5	1;100
1;50	25;10	25;5	50;5
5;25	5;50		

# Information Retrieval Style Evaluation

Evaluated 10 systems ( $iter;m$ ) with different  $iter$  and  $m$

10;10	10;25	10;5	1;100
1;50	25;10	25;5	50;5
5;25	5;50		

10;25
5;50



# Information Retrieval Style Evaluation

Evaluated 10 systems ( $iter;m$ ) with different  $iter$  and  $m$

10;10	10;25	10;5	1;100
1;50	25;10	25;5	50;5
5;25	5;50		

10;25
5;50



## ■ Results on H2 annotations

System	Pr	Rec	F1
10;25	.53	.71	.60
5;50	.54	.72	.62



# Information Retrieval Style Evaluation

Evaluated 10 systems ( $iter;m$ ) with different  $iter$  and  $m$

10;10	10;25	10;5	1;100
1;50	25;10	25;5	50;5
5;25	5;50		

10;25
5;50



## ■ Results on H2 annotations

System	Pr	Rec	F1
10;25	.53	.71	.60
5;50	.54	.72	.62

## ■ Results on H1 annotations

System	Pr	Rec	F1
10;25	.46	.85	.59
5;50	.46	.84	.58

# Visual or Non-Visual



Visual

# Visual or Non-Visual



Visual

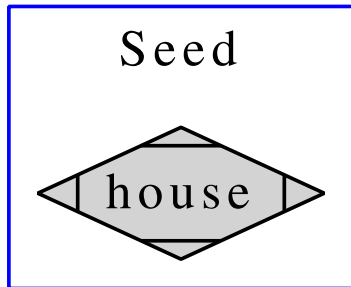
Non Visual

- Graph is constructed
  - Nodes: Nouns
  - Neighbors: Top 10 similar nouns
  - Edges: Distributional similarity as weights
- Seeds
  - Example **visual: car house tree horse bottle**
- In-Degree (inDgr) [Kozareva et al., 2008]
- Mutual Exclusion [Thelen and Riloff, 2002]

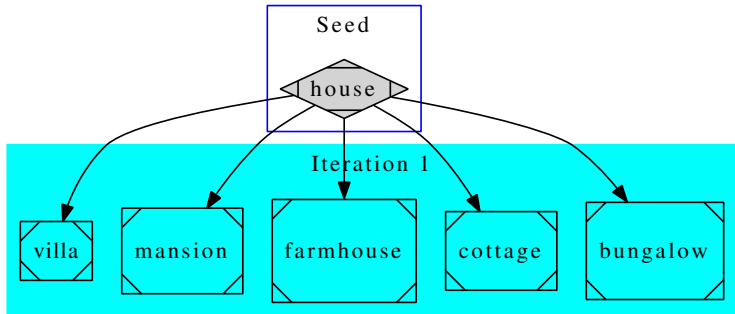
# Visual and Non visual seeds for nouns and adjectives

<b>Nouns</b>		<b>Adjectives</b>	
Visual	Non Visual	Visual	Non Visual
car	idea	brown	public
house	bravery	green	original
tree	deceit	wooden	whole
horse	trust	shiny	initial
animal	dedication	rusty	total
man	anger	rectangular	personal
table	humour	furry	intrinsic
bottle	luck	striped	individual
woman	inflation	orange	political
computer	honesty	feathered	righteous

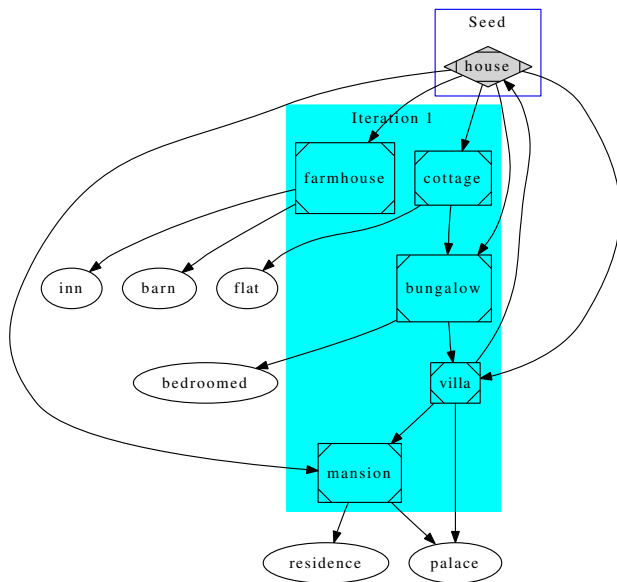
# Bootstrapping in Action



# Bootstrapping in Action



# Bootstrapping in Action





## What % of flicker data-set nouns are visual and non-visual?

- Visual nouns: 64.23%
- Non Visual nouns: 17.71

## **What % of flicker data-set nouns are visual and non-visual?**

- Visual nouns: 64.23%
- Non Visual nouns: 17.71

## **What % of flicker data-set adjectives are visual and non-visual?**

- Visual adjectives: 51.79%
- Non Visual adjectives: 14.40%

## How many visual nouns are physical entities in word net?

- Physical: 49.17%
- Abstract: 19.81%
- Unknown: 31.02%

## How many visual nouns are physical entities in word net?

- Physical: 49.17%
- Abstract: 19.81%
- Unknown: 31.02%

## How many non-visual nouns are abstract entities in word net?

- Physical: 25.30%
- Abstract: 42.15%
- Unknown: 32.56%