DEcompositional Distributional Semantics

Marco Baroni

Center for Mind/Brain Sciences
University of Trento

CIS Seminar
Munich
October 21st 2014
Acknowledging...

Georgiana Dinu  Angeliki Lazaridou

COMPOSES: COMPositional Operations in SEmantic Space
Outline

(Compositional) distributional semantics

DEcompositional distributional semantics

Linguistic evaluation

Cross-modal evaluation
The distributional hypothesis
Harris, Charles and Miller, Firth, Wittgenstein? . . .

The meaning of a word is (can be approximated by, learned from) the set of contexts in which it occurs in texts

We found a little, hairy \textit{wampimuk} sleeping behind the tree

See also MacDonald & Ramscar CogSci 2001
he curtains open and the moon shining in on the barely
ars and the cold, close moon ". And neither of the w
rough the night with the moon shining so brightly, it
made in the light of the moon. It all boils down, wr
surely under a crescent moon, thrilled by ice-white
sun, the seasons of the moon? Home, alone, Jay pla
m is dazzling snow, the moon has risen full and cold
un and the temple of the moon, driving out of the hug
in the dark and now the moon rises, full and amber a
bird on the shape of the moon over the trees in front
But I could n't see the moon or the stars, only the
orning, with a sliver of moon hanging among the stars
they love the sun, the moon and the stars. None of
the light of an enormous moon. The plash of flowing w
man 's first step on the moon; various exhibits, aer
the inevitable piece of moon rock. Housing The Airsh
oud obscured part of the moon. The Allied guns behind
# Distributional semantics

Distributional meaning encoded in (functions-of-)co-occurrence vectors

<table>
<thead>
<tr>
<th></th>
<th>planet</th>
<th>night</th>
<th>full</th>
<th>shadow</th>
<th>shine</th>
<th>crescent</th>
</tr>
</thead>
<tbody>
<tr>
<td>moon</td>
<td>10</td>
<td>22</td>
<td>43</td>
<td>16</td>
<td>29</td>
<td>12</td>
</tr>
<tr>
<td>sun</td>
<td>14</td>
<td>10</td>
<td>4</td>
<td>15</td>
<td>45</td>
<td>0</td>
</tr>
<tr>
<td>dog</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Word embeddings
A representation learning alternative to traditional context vectors

... see millions of stars on a dark night

Bengio et al. JMLR 2003, Collobert and Weston ICML 2008, Turian et al. ACL 2010, Mikolov et al. NIPS 2013, ..., systematic comparison in Baroni et al. ACL 2014
Distributional semantics
The geometry of meaning

<table>
<thead>
<tr>
<th></th>
<th>shadow</th>
<th>shine</th>
</tr>
</thead>
<tbody>
<tr>
<td>moon</td>
<td>16</td>
<td>29</td>
</tr>
<tr>
<td>sun</td>
<td>15</td>
<td>45</td>
</tr>
<tr>
<td>dog</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

sun (15,45)
moon (16,29)
dog (10,0)
Geometric neighbours \(\approx\) semantic neighbours

<table>
<thead>
<tr>
<th>rhino</th>
<th>fall</th>
<th>good</th>
<th>sing</th>
</tr>
</thead>
<tbody>
<tr>
<td>woodpecker</td>
<td>rise</td>
<td>bad</td>
<td>dance</td>
</tr>
<tr>
<td>rhinoceros</td>
<td>increase</td>
<td>excellent</td>
<td>whistle</td>
</tr>
<tr>
<td>swan</td>
<td>fluctuation</td>
<td>superb</td>
<td>mime</td>
</tr>
<tr>
<td>whale</td>
<td>drop</td>
<td>poor</td>
<td>shout</td>
</tr>
<tr>
<td>ivory</td>
<td>decrease</td>
<td>improved</td>
<td>sound</td>
</tr>
<tr>
<td>plover</td>
<td>reduction</td>
<td>perfect</td>
<td>listen</td>
</tr>
<tr>
<td>elephant</td>
<td>logarithm</td>
<td>clever</td>
<td>recite</td>
</tr>
<tr>
<td>bear</td>
<td>decline</td>
<td>terrific</td>
<td>play</td>
</tr>
<tr>
<td>satin</td>
<td>cut</td>
<td>lucky</td>
<td>hear</td>
</tr>
<tr>
<td>sweatshirt</td>
<td>hike</td>
<td>smashing</td>
<td>hiss</td>
</tr>
</tbody>
</table>
Distributional semantics: A general-purpose representation of lexical meaning
Baroni and Lenci 2010

- Similarity (\textit{cord-string} vs. \textit{cord-smile})
- Synonymy (\textit{zenith-pinnacle})
- Concept categorization (\textit{car ISA vehicle}; \textit{banana ISA fruit})
- Selectional preferences (\textit{eat topinambur} vs. \textit{*eat sympathy})
- Analogy (\textit{mason} is to \textit{stone} like \textit{carpenter} is to \textit{wood})
- Relation classification (\textit{exam-anxiety} are in \textsc{cause-effect} relation)
- Qualia (TELIC ROLE of \textit{novel} is \textit{to entertain})
- Salient properties (\textit{car-wheels}; \textit{dog-barking})
- Argument alternations (\textit{John broke the vase} - \textit{the vase broke}, \textit{John minces the meat} - \textit{*the meat minced})
To kill . . .

<table>
<thead>
<tr>
<th>object</th>
<th>cosine</th>
<th>with</th>
<th>cosine</th>
</tr>
</thead>
<tbody>
<tr>
<td>kangaroo</td>
<td>0.51</td>
<td>hammer</td>
<td>0.26</td>
</tr>
<tr>
<td>person</td>
<td>0.45</td>
<td>stone</td>
<td>0.25</td>
</tr>
<tr>
<td>robot</td>
<td>0.15</td>
<td>brick</td>
<td>0.18</td>
</tr>
<tr>
<td>hate</td>
<td>0.11</td>
<td>smile</td>
<td>0.15</td>
</tr>
<tr>
<td>flower</td>
<td>0.11</td>
<td>flower</td>
<td>0.12</td>
</tr>
<tr>
<td>stone</td>
<td>0.05</td>
<td>antibiotic</td>
<td>0.12</td>
</tr>
<tr>
<td>fun</td>
<td>0.05</td>
<td>person</td>
<td>0.12</td>
</tr>
<tr>
<td>book</td>
<td>0.04</td>
<td>heroin</td>
<td>0.12</td>
</tr>
<tr>
<td>conversation</td>
<td>0.03</td>
<td>kindness</td>
<td>0.07</td>
</tr>
<tr>
<td>sympathy</td>
<td>0.01</td>
<td>graduation</td>
<td>0.04</td>
</tr>
</tbody>
</table>
The infinity of sentence meaning

I know a mouse. and he hasn’t got a house. Who put all those things in your hair.

The black and green scarecrow is sadder than me.

I want to tell you a story About a little man if I can. Let’s go into the other room and make some tea.

Eleanor Rigby died in the church and was buried along with her name. Eating, sleeping, drinking their wine. Someone is speaking but she doesn’t know.

Eleanor Rigby picks up the rice in the church where a wedding has been. She’s so near.

Oh Mother, tell me more. You can’t see me But I can you.

Yes they did. Because I’m the taxman. yeah, I’m the taxman.

I’m in bed Achin’ head Cold is lead Choke on bread Under fed Gold is lead Jesus bled Pain is red Are gone Grow Grassy spoon You swoon June bloom Mussin.

The seven is the number of the young light. Blinding signs flap, Flicker, flicker, flicker blam. He does everything he can. Doctor Robert.

Action brings good fortune. So it is not dying. All the lonely people it forms when darkness.

That cat’s something I can’t explain. All the lonely people Where do they all come from?

He didn’t care. You anything, everything if you want things.

Good Day Sunshine. Looking for the words of a sermon that no one will hear.

I’m in bed Achin’ head Cold is lead Choke on bread Under fed Gold is lead Jesus bled Pain is red Are gone Grow Grassy spoon You swoon June bloom Mussin.

You don’t understand what I said. Father McKenzie writing the words of a sermon that no one will hear.

We all live in our yellow submarine Nobody can deny that there’s something there.

Keeping an eye on the world going by my window.

When I was a boy everything was right. Another father McKenzie writing the words of a sermon that no one will hear.

Yippee! Father McKenzie writing the words of a sermon that no one will hear.

I was a king who ruled the land. Know what it is to be sad… Good Day Sunshine.

The black and green scarecrow as everyone knows stood. I don’t mind. I think they’re crazy. Doctor kindly tell your wife that I’m alive.

Do you know Mrs. Brown has been a librarian. Take a drink from his special cup. Doctor Robert.

He’s getting rather old. but he’s a good mouse. There, running my hands through his

And then one day - hooray! Taking my time. From time to time and limb green. a second scene A flight between the blue you once knew.

Why’dya have to leave me there Hangin’ in my infant air Waiting? Should five per cent appear too small. Be thankful I don’t take it all.

I want her everywhere things that make me feel that I’m mad. You tell me that you’ve got everything you want And your bird can sing.

You tell me that you’ve heard every sound there is And your bird can swim. Who is it for? Jupiter and

Don’t pay money just to see yourself with Doctor. If you’re down hell pick you up. Doctor Robert.

But now he’s resigned to his Lying there and staring at the ceiling. If you don’t want to pay some more. I’ve gone.

Leaving me where I am. What does he care? Dewberry’s socks in the

Blind and limpid green The sounds surround the icy waters underground Lime and limpid green The sounds surround the icy waters underground. Waiting for a sleepy feeling. Please. don’t spoil my day. I’m miles away. And all I’m only sleeping.

No, no. no. your wrong. With silver eyes. In the town.

That's something I can’t explain. All the lonely people Where do they all come from?

Wandering and dreaming The words have different meaning. I feel good in a special way. Because you’re making me feel.

Good Day Sunshine. You’re the kind of girl that fits in with my world.

I was a boy everything was right. Another father McKenzie writing the words of a sermon that no one will hear.

Good Day Sunshine. For you anything, everything if you want things.

Now what is it to be sad. Running everywhere at such a speed Till they find there’s no need. And you’re making me feel like I’ve never been born.

And she’s making me feel like I’ve never been born. He wore a scarlet tunic. A blue green hood. It looked quite good. Take a drink from his special cup. Doctor Robert.

Some rhyme. Some thing. Most of them are clockwork.
Compositionality
The meaning of an utterance is a function of the meaning of its parts and their composition rules (Frege 1892)
Compositional distributional semantics: What for?

NOT this:

every farmer who owns a donkey beats it

∀x∀y(FARMER(x) ∧ DONKEY(y) ∧ OWNS(x, y) → BEAT(x, y))
Compositional distributional semantics: What for?

Word meaning in context (Mitchell and Lapata ACL 2008)

Paraphrase detection (Blacoe and Lapata EMNLP 2012)

the cucumber is rotten

the cucumber is old

the cucumber is ancient

"cookie dwarfs hop under the crimson planet"

"gingerbread gnomes dance under the red moon"

"red gnomes love gingerbread cookies"

"students eat cup noodles"
Compositional distributional semantics: How?

From:
Simple functions

\[ \text{very} \quad \rightarrow \quad + \quad \rightarrow \quad \text{good} \quad \rightarrow \quad + \quad \rightarrow \quad \text{movie} \]

\[ \text{very good movie} \]

Mitchell and Lapata
ACL 2008

To:
Complex composition operations

Socher at al. EMNLP 2013
General estimation of compositional models
Dinu, Pham and Baroni CVSC 2013; also: Guevara GEMS 2010, Baroni and Zamparelli EMNLP 2010

Use (reasonably frequent) corpus-extracted phrase vectors to learn the parameters of composition functions:

\[ \theta^* = \text{argmin}_{\theta} ||P - f_{\text{comp}}(U, V)||^2 \]

\( P/U, V \) - Phrase/Input occurrence matrices
General estimation of compositional models

\[
\begin{align*}
[\text{white} \rightarrow \text{man}] & \rightarrow \text{white.man} \\
[\text{red} \rightarrow \text{car}] & \rightarrow \text{red.car} \\
[\text{large} \rightarrow \text{man}] & \rightarrow \text{large.man} \\
[\text{red} \rightarrow \text{dress}] & \rightarrow \text{red.dress} \\
[\text{white} \rightarrow \text{wine}] & \rightarrow \text{white.wine}
\end{align*}
\]
The linear Full Additive composition model
Guevara GEMS 2010, Zanzotto et al. COLING 2010

- Given two word vectors $\vec{u}$ and $\vec{v}$ in syntactic relation $R$
  compute phrase vector $\vec{p}$:

$$\vec{p} = A_R \vec{u} + B_R \vec{v} = [A_R, B_R] \begin{bmatrix} \vec{u} \\ \vec{v} \end{bmatrix}$$

- Parameters: syntax-dependent matrices $A_R$ and $B_R$

- General estimation from corpus-extracted phrase and word vectors as least-squares regression problem:

$$\arg\min_{A_R, B_R} ||P - [A_R, B_R] \begin{bmatrix} U \\ V \end{bmatrix}||^2$$
Outline

(Compositional) distributional semantics

DEcompositional distributional semantics

Linguistic evaluation

Cross-modal evaluation
Generation in distributional semantics

Dinu and Baroni ACL 2014

gingerbread gnomes dance under the red moon

gingerbread gnomes
dance under the red moon

...
What for?

- Theoretical: linguistic communication goes both ways: understanding (from words to meaning) but also production (from meaning to words)

- Practical: Can’t search through infinite phrase/sentence space to find semantic neighbours of a target item!

  \[ \Rightarrow \] Current compositional distributional semantics limited to “static” tasks where list of items of interest can be precompiled in advance

- Applications:
  - True paraphrase generation
  - Compose phrase vector in language X, decode in language Y
  - From different modalities to language:
    - Generate a verbal description from an image-representing vector
    - Spell out a thought encoded in a brain activation vector
Generation through decomposition functions

1. Decomposition
\[
[\vec{u}; \vec{v}] = f_{\text{decomp}}(\vec{p})
\]

\(f_{\text{decomp}} : \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{R}^d\)

\(\vec{p} \in \mathbb{R}^d\) phrase vector
\(\vec{u}, \vec{v} \in \mathbb{R}^d\) child vectors

2. Nearest neighbor queries
\(word = \text{NN}_{\text{Lex}}(\vec{u})\)
Estimating decomposition functions from corpus-observed phrase vectors

Example: $f_{\text{decomp}_{\text{AN}}}$

- $\overrightarrow{\text{white} \cdot \text{man}} \rightarrow [\overrightarrow{\text{white}}; \overrightarrow{\text{man}}]$
- $\overrightarrow{\text{red} \cdot \text{car}} \rightarrow [\overrightarrow{\text{red}}; \overrightarrow{\text{car}}]$
- $\overrightarrow{\text{large} \cdot \text{man}} \rightarrow [\overrightarrow{\text{large}}; \overrightarrow{\text{man}}]$
- $\overrightarrow{\text{red} \cdot \text{dress}} \rightarrow [\overrightarrow{\text{red}}; \overrightarrow{\text{dress}}]$
- $\overrightarrow{\text{white} \cdot \text{wine}} \rightarrow [\overrightarrow{\text{white}}; \overrightarrow{\text{wine}}]$
Learning linear decomposition functions

- Estimate syntax-specific linear transformations $W_R \in \mathbb{R}^{2d \times d}$:

  $$f_{\text{decomp}_R}(\bar{p}) = W_R \bar{p}$$
  $$\arg\min_{W_R \in \mathbb{R}^{2d \times d}} \| [U; V] - W_R P \|^2$$

- Corpus: Wikipedia + BNC + ukWaC (2.8 billion words) (Italian: itWaC, 1.8 billion words)
- Use vectors of top 50K most frequent phrase-child1-child2 tuples for decomposition function training
- Word/corpus-based phrase vectors: created with word2vec: \textit{cbow} method, 300 dimensions, 5-word context windows, negative sampling
Outline

(Compositional) distributional semantics

DEcompositional distributional semantics

Linguistic evaluation

Cross-modal evaluation
Paraphrasing

Noun to Adj-Noun paraphrasing
620 pairs, adapted from Turney JAIR 2012:

fallacy    false belief
charisma   personal appeal
binary     double star

Adj-Noun to Noun-Prep-Noun dataset
192 pairs:

pre-election promises    promises before election
inexperienced user       user without experience
Paraphrase generation schemes

Noun to Adj-Noun
Decompose N into AN

\[ f_{\text{decomp}_{AN}} \text{ fallacy} \rightarrow [\text{false} \ ; \ \text{belief}] \]

Adj-Noun to Noun-Prep-Noun:
Compose AN - decompose AN - decompose A into PN

\[ f_{\text{comp}_{AN}} [\text{pre-election} \ ; \ \text{promises}] \rightarrow \bar{p} \]
\[ f_{\text{decomp}_{AN}} \bar{p} \rightarrow [\bar{u} \ ; \ \text{promises}] \]
\[ f_{\text{decomp}_{PN}} \bar{u} \rightarrow [\text{before} \ ; \ \text{election}] \]
Results

Search space: 20,000 Adjs and 20,000 Nouns, 14 Prepositions

### N-AN

<table>
<thead>
<tr>
<th></th>
<th>Adjective</th>
<th></th>
<th>Noun</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Direct</td>
<td>Decompose</td>
<td>Direct</td>
<td>Decompose</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.15</td>
<td>0.10</td>
<td>0.06</td>
<td>0.12</td>
</tr>
<tr>
<td>Median rank</td>
<td>168</td>
<td>67</td>
<td>60</td>
<td>29</td>
</tr>
</tbody>
</table>

### AN-NPN

<table>
<thead>
<tr>
<th></th>
<th>Noun</th>
<th>Prep</th>
<th>Noun</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Direct</td>
<td>Decompose</td>
<td>Direct</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.93</td>
<td>0.98</td>
<td>0.15</td>
</tr>
<tr>
<td>Median rank</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
</tbody>
</table>
## N-AN examples

<table>
<thead>
<tr>
<th>Noun</th>
<th>Adj Noun</th>
<th>Gold</th>
</tr>
</thead>
<tbody>
<tr>
<td>thunderstorm</td>
<td>thundery storm</td>
<td>electrical storm</td>
</tr>
<tr>
<td>reasoning</td>
<td>deductive thinking</td>
<td>abstract thought</td>
</tr>
<tr>
<td>jurisdiction</td>
<td>legal authority</td>
<td>legal power</td>
</tr>
<tr>
<td>folk</td>
<td>local music</td>
<td>common people</td>
</tr>
<tr>
<td>vitriol</td>
<td>political bitterness</td>
<td>sulfuric acid</td>
</tr>
<tr>
<td>superstition</td>
<td>old-fashioned religion</td>
<td>superstitious notion</td>
</tr>
<tr>
<td>zoom</td>
<td>fantastic camera</td>
<td>rapid growth</td>
</tr>
<tr>
<td>Adj Noun</td>
<td>Noun Prep Noun</td>
<td>Gold</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------------------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>mountainous region</td>
<td>region in highlands</td>
<td>region with mountains</td>
</tr>
<tr>
<td>undersea cable</td>
<td>cable through ocean</td>
<td>cable under sea</td>
</tr>
<tr>
<td>inter-war years</td>
<td>years during 1930s</td>
<td>years between wars</td>
</tr>
<tr>
<td>post-operative pain</td>
<td>pain through patient</td>
<td>pain after operation</td>
</tr>
<tr>
<td>superficial level</td>
<td>level between levels</td>
<td>level on surface</td>
</tr>
</tbody>
</table>
Cross-lingually

Use 5K English-Italian word pairs to estimate linear functions mapping word vectors from one language onto the other (Mikolov, Le and Sutskever arXiv 2013)

<table>
<thead>
<tr>
<th>English</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>car</td>
<td></td>
</tr>
<tr>
<td>news</td>
<td></td>
</tr>
<tr>
<td>world</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Italian</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>auto</td>
<td></td>
</tr>
<tr>
<td>notizie</td>
<td></td>
</tr>
<tr>
<td>mondo</td>
<td></td>
</tr>
</tbody>
</table>
Testing on 1,000 AN pairs from a movie subtitles phrase table:

- spectacular woman ➔ donna affascinante
- vicious killer ➔ killer pericoloso
Results

- Chance accuracy: $\frac{1}{20,000^2}$
- Using generation confidence (cosine to nearest neighbour) as threshold

<table>
<thead>
<tr>
<th>Threshold</th>
<th>En→It</th>
<th>Coverage</th>
<th>It→En</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.21</td>
<td>100%</td>
<td>0.32</td>
<td>100%</td>
</tr>
<tr>
<td>0.55</td>
<td>0.25</td>
<td>70%</td>
<td>0.40</td>
<td>63%</td>
</tr>
<tr>
<td>0.60</td>
<td>0.31</td>
<td>32%</td>
<td>0.45</td>
<td>37%</td>
</tr>
<tr>
<td>0.65</td>
<td>0.45</td>
<td>9%</td>
<td>0.52</td>
<td>16%</td>
</tr>
<tr>
<td>English</td>
<td>Italian</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>----------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>black tie</td>
<td>cravatta nera</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>indissoluble tie</td>
<td>alleanza indissolubile</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vicious killer</td>
<td>assassino feroce (killer pericoloso)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rough neighborhood</td>
<td>zona malfamata (ill-repute zone, quartiere difficile)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mortal sin</td>
<td>peccato eterno (eternal sin)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>canine star</td>
<td>stella stellare (stellar star)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Outline

(Compositional) distributional semantics

DEcompositional distributional semantics

Linguistic evaluation

Cross-modal evaluation
Object recognition in computer vision

Is it a CAR?

Is it a BOAT?

CAT!

http://abstrusegoose.com/496
Vector representation of images
The Bags-of-Visual-Words (BOVW) pipeline in image analysis
(Sivic and Zisserman ICCV 2003)
Learning a vector-based mapping of visual objects to words
Mapping pictures of previously unseen objects onto linguistic space

Frome et al. NIPS 2013, Socher et al. NIPS 2013, Lazaridou et al. ACL 2014 (“zero-shot learning”)
Inducing the cross-modal mapping function

cat → text-based vector
minimize, e.g., cosine
translated vector

REGRESSION ALGORITHM

e.g., ridge regression
BoVW vector
Images depict **visual phrases**

Lazaridou, Dinu, Liška and Baroni *submitted*
Images depict *visual phrases*

*Direct* mapping

- Map image vector onto linguistic space
- Retrieve noun with nearest vector to label image

**Decompositional approach (**Decompose**)**

- Map image vector onto linguistic space
- DEcompose mapped vector with AN decomposition function
- Retrieve adjective with vector nearest to A component and noun with vector nearest to N component
From images to phrases

DECOMPOSE

MAP

orange

liqueur
Gory details

- Image vectors: PHOW-color-based Bag-of-Visual-Words features, pyramid representation, $600 \times 20$ dimensions

- Cross-modal mapping learned by linear regression from about 23K ImageNet pictures representing 750 objects labeled with nouns (not in test set)
  - NB: no adjectives, no phrases seen in cross-modal mapping training

- Same adjective-noun decomposition function as in experiments above
  - NB: no visual data used to train decomposition function
Test set
Adapted from Russakovsky and Fei-Fei ECCV 2010

- 8.5K annotated images, annotated for objects-nouns and attributes-adjectives
- 203 distinct objects, 25 attributes (2.7 on average per-image)

- furry, white, smooth cat

- green, shiny cocktail
Results
Percentages across images

Search space: 5,000 Adjectives, 10,000 Nouns

<table>
<thead>
<tr>
<th></th>
<th>Direct</th>
<th></th>
<th></th>
<th>Decompose</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A hit</td>
<td>A recall</td>
<td>N hit</td>
<td>A hit</td>
<td>A recall</td>
<td>N hit</td>
</tr>
<tr>
<td>@1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>@5</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>31</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>@10</td>
<td>9</td>
<td>4</td>
<td>5</td>
<td>44</td>
<td>23</td>
<td>14</td>
</tr>
<tr>
<td>@20</td>
<td>19</td>
<td>9</td>
<td>9</td>
<td>59</td>
<td>35</td>
<td>20</td>
</tr>
<tr>
<td>@50</td>
<td>43</td>
<td>22</td>
<td>20</td>
<td>81</td>
<td>59</td>
<td>29</td>
</tr>
<tr>
<td>@100</td>
<td>67</td>
<td>44</td>
<td>33</td>
<td>89</td>
<td>70</td>
<td>41</td>
</tr>
</tbody>
</table>
Examples

GOLD: white, brown dog
Decompose: white dog
Direct: animal goat

GOLD: shiny, round syrup
Decompose: shiny flan
Direct: crunchy ramekin
Using Decomposition for attribute-based object classification

thick, wet, dry, cylindrical, motionless, translucent aeroplane

cuddly, wild, cute, furry, white, coloured dog

Adding Decompose-generated adjective features ("attributes") to a standard BoVW-classifier brings object-recognition up from 30% to 36% on subset of VOC 2008
What next?

- More powerful decomposition functions, more training signals
- Learn semantic and syntactic decompositions jointly
- Generate longer phrases, sentences
- More targeted evaluation sets
- Make phrase generation more human-like
  - From *true* to *salient* (*metal car* vs. *red car*)
  - Consider linguistic naturalness of phrases (*damp lips* vs. *moist lips*)
  - Take more inter-dependencies into account (if *aeroplane* is *wet* it cannot be *dry*)
- Associate multiple objects in image with phrases of different syntactic structures
That’s all, folks!

Thank you!

http://clic.cimec.unitn.it/marco