

Mining Semantic Affordances of Visual Object Categories

Yu-Wei Chao, Zhan Wang, Rada Mihalcea, Jia Deng

Computer Science & Engineering, University of Michigan, Ann Arbor



Introduction

- We introduce the problem of mining the knowledge of **semantic affordance**: given an action and an object, determine whether the action can be applied to the object.

Why useful?

- It gives a list of valid action-object pairs, the basis for creating a large-scale image/video dataset to train and evaluate action recognition algorithms.
- Affordance knowledge can be used as a common sense prior for generating natural language descriptions from images or videos.

Contributions

- Introduce the new problem of mining semantic affordances.
- A benchmark dataset for affordance modeling that contains the complete ground truth for all 20 PASCAL VOC categories on 957 verb synsets.
- Explore and analyze a variety of approaches and present a number of significant insights.

Problem Formulation

Question : How do we obtain candidate action/object categories?

- The same word can have different meanings.

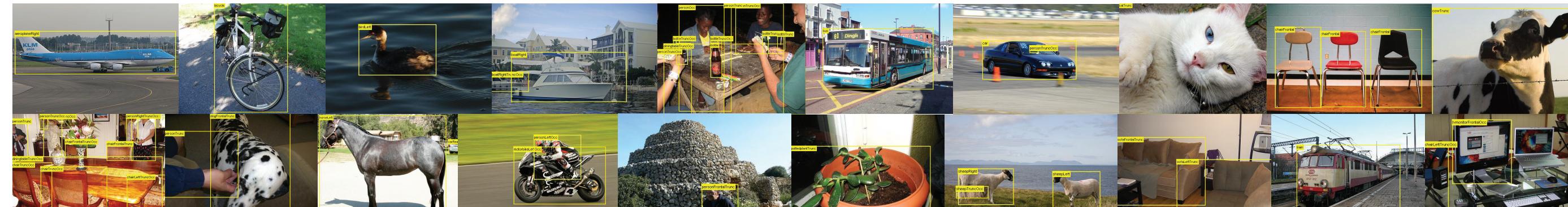
draw 1: represent by making a drawing of, as with a pencil, chalk, etc. on a surface
Draw me a horse.

2: take liquid out of a container or well
She drew water from the barrel

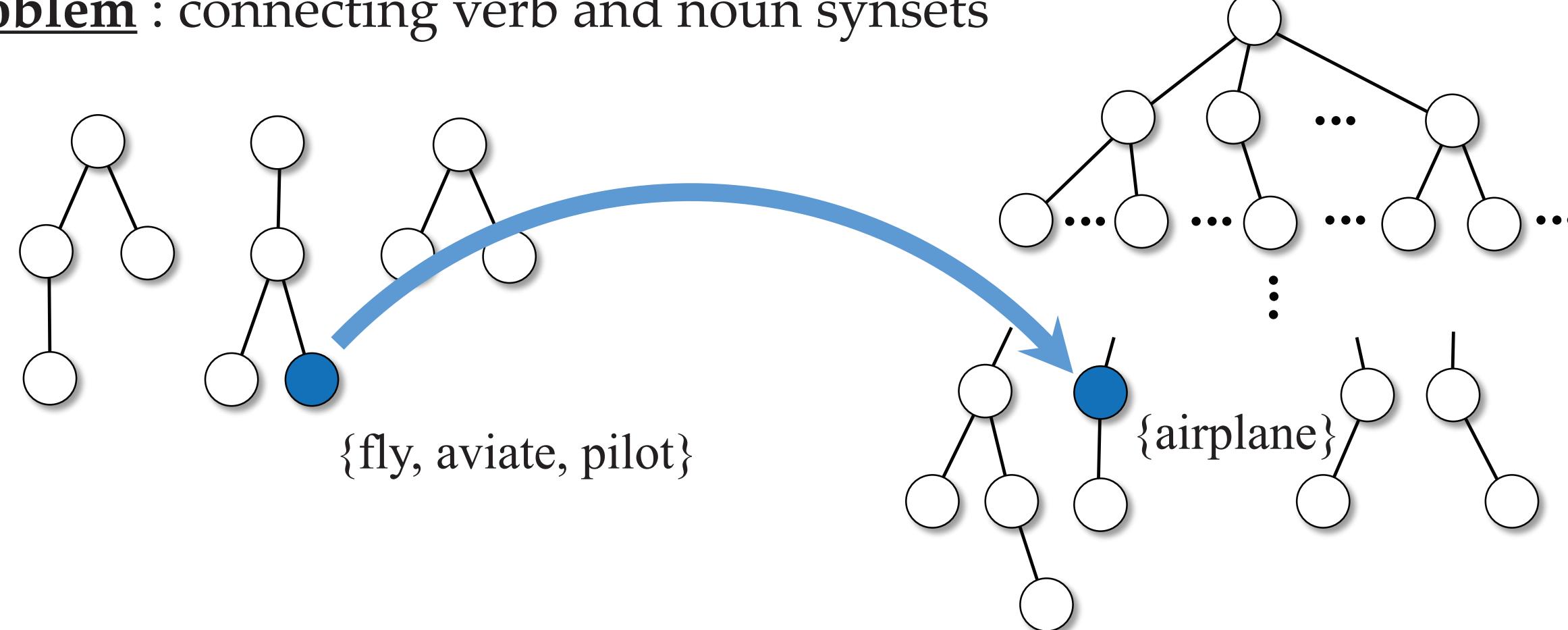
Wordnet — verb and noun synsets

{wash, launder}
{capture, catch}
{draw, take out}
{switch off, cut, turn off, turn out}

For noun synsets, we use 20 objects classes from PASCAL VOC



Problem : connecting verb and noun synsets



Crowdsourcing Semantic Affordances

- Selecting verb synsets — Wordnet has 13,767 verb synsets. We are interested in the verb synsets that are both **common** and **visual**.

1. Common : Google Syntactic N-grams dataset

drive **drive/VB/ROOT/0** **car/NN/** **dobj/1** **21733**

2. Visual : Crowdsourcing using Amazon Mechanical Turk (AMT)

5. align

definition: place in a line or arrange so as to be parallel or straight
synonyms: align, line_up, adjust
example: align the car with the curb

Is it possible to tell whether someone is "**align** -ing" something by looking at an image or watching a video clip without sound?

Visualness	Synset Synonyms	Definition
Definitely yes	{wash, launder} {drive}	Cleanse with a cleaning agent, such as soap, and water. Operate or control a vehicle.
Yes	{deliver} {switch, off, cut, turn off, turn out}	Bring to a destination, make a delivery. Cause to stop operating by disengaging a switch.
Maybe	{enjoy} {allow, grant}	Have for one's benefit. Let have.
No	{develop, acquire, evolve} {drive}	Gain through experience. Compel somebody to do something, often against his own will or judgment.
Definitely no or Make no sense	{wish} {come} {decline, go down, wane}	Make or express a wish. Come to pass, arrive, as in due course. Grow smaller.

Crowdsourcing semantic affordances

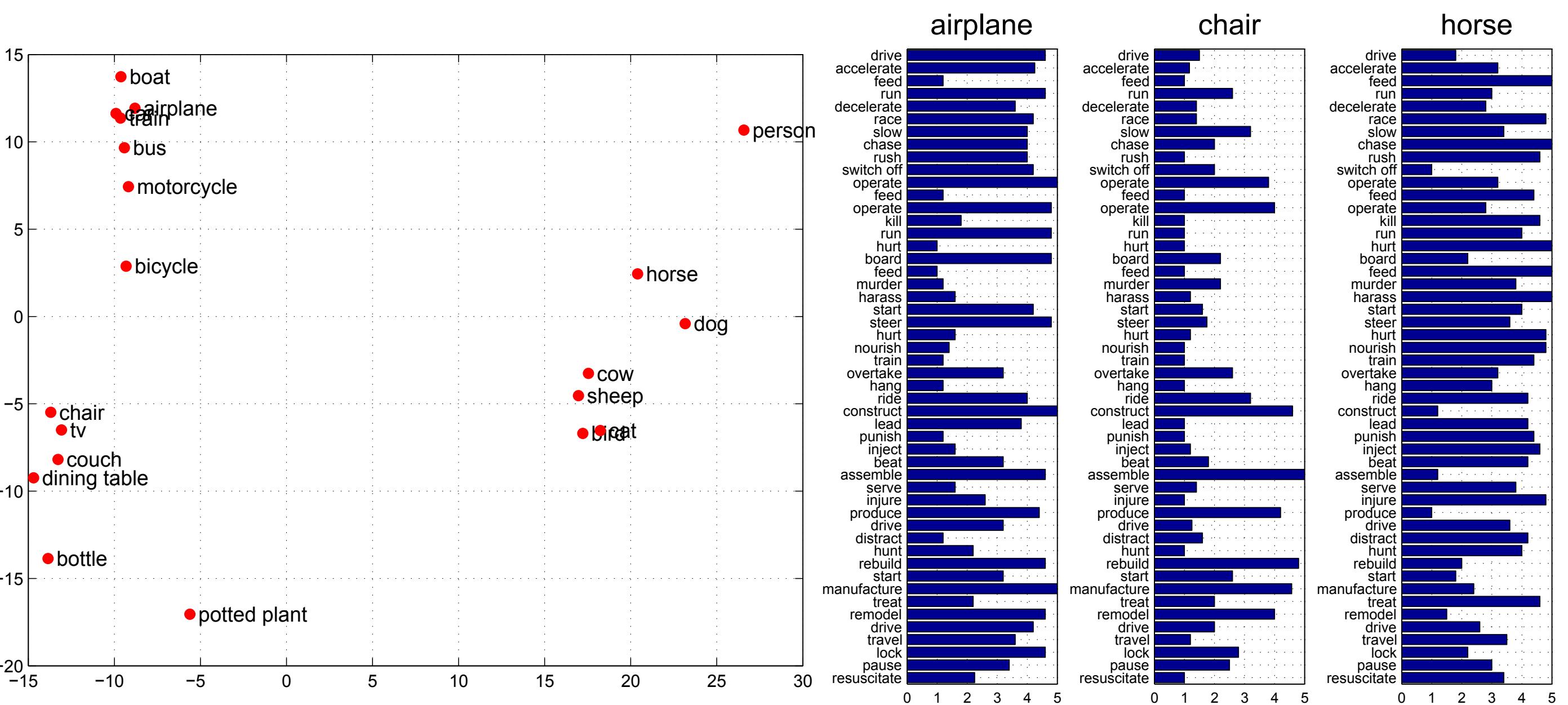
20 noun synsets \times ~1000 verb synsets \times 5 Turkers \approx 100K questions

1. load

definition: fill or place a load on
synonyms: load, lade, laden, load_up
example: load a car

Is it possible to **load** (fill or place a load on) a **car**?

Plausibility	Action		Object
	Synset Synonyms	Definition	Synset Synonyms
Definitely yes	{race, run}	Compete in a race.	{car, auto, automobile, machine, motorcar}
Yes	{turn}	Cause to move around or rotate.	{sofa, couch, lounge}
Maybe	{compress, constrict, squeeze, compact, contract, press}	Squeeze or press together.	{bottle}
No	{capture, catch}	Capture as if by hunting, snaring or trapping.	{chair}
Definitely no or Make no sense	{wear, bear} {cultivate, crop, work}	Have on one's person. Prepare for crops.	{airplane, aeroplane, plane} {person, individual, somebody, mortal, soul}

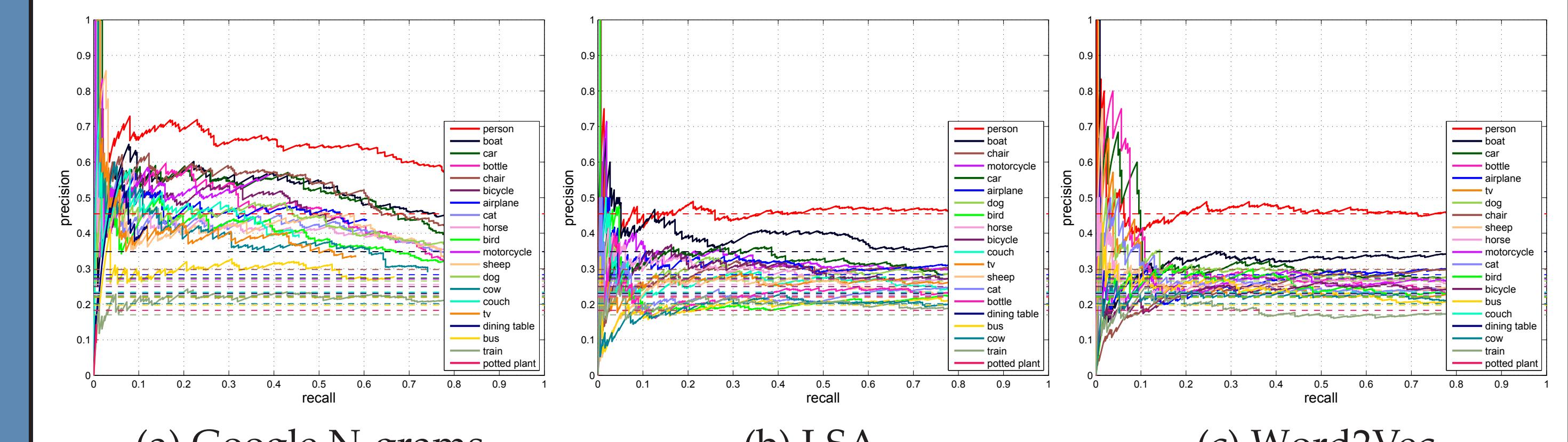


Mining Semantic Affordances

- Question** : To what extent can we automatically extract semantic affordances?

- Binary classification problem (using crowdsourced ground truth)

1. Mining from texts — co-occurrence information of verbs and nouns



(a) Google N-grams

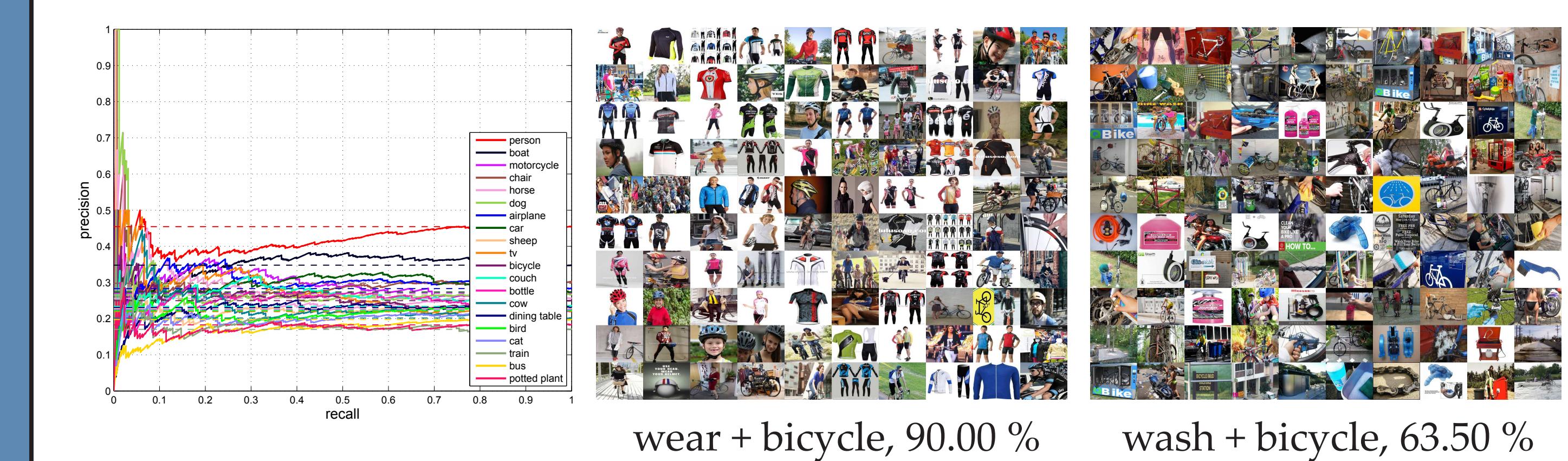
(b) LSA

(c) Word2Vec

Google N-Gram	Action		Object
	Synset Synonyms	Definition	Synset Synonyms
True positives	{draw}	Represent by making a drawing of, as with a pencil, chalk, etc. on a surface.	{person, individual, somebody, mortal, soul}
	{pass, hand, reach, pass on, turn over, give}	Place into the hands or custody of.	{bottle}
False positives	{draw, take out}	Take liquid out of a container or well.	{person, individual, somebody, mortal, soul}
	{pass, go through, go across}	Go across or through.	{bottle}
False negatives	{photograph, snap, shoot}	Record on photographic film.	{airplane, aeroplane, plane}
	{award, present}	Give, especially as an honor or reward.	{person, individual, somebody, mortal, soul}

2. Mining from images

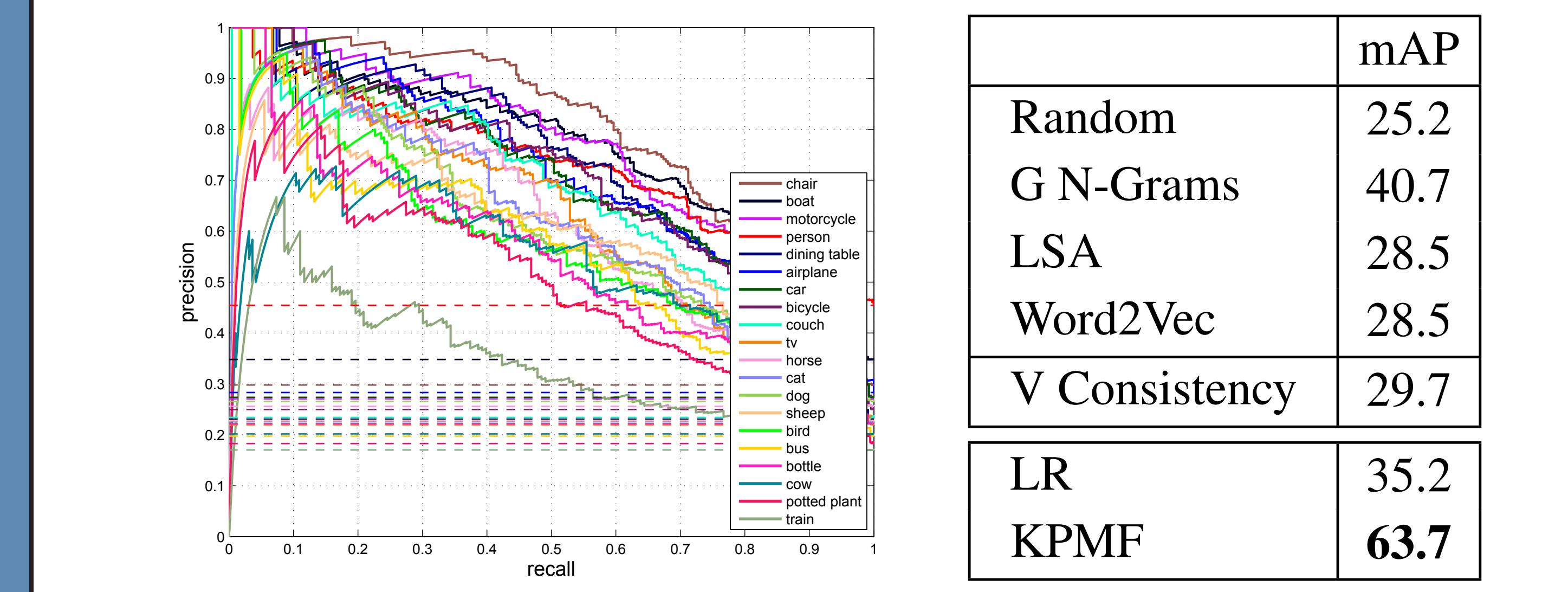
- Query verb-noun pairs in Google Image Search
- Measure **visual consistency** by SVM cross validation accuracy



wear + bicycle, 90.00 %
wash + bicycle, 63.50 %

3. Collaborative filtering

- User vs. movie \longrightarrow object vs. action
- Predict if an object "likes" an action
- Kernelized Probabilistic Matrix Factorization (KPMF)
- Leave-one-class-out prediction



	mAP
Random	25.2
G N-Grams	40.7
LSA	28.5
Word2Vec	28.5
V Consistency	29.7
LR	35.2
KPMF	63.7