Aalto University
School of Science

META-NET Workshop in ICANN 2011: Context in Machine Translation

Visual context for natural language processing

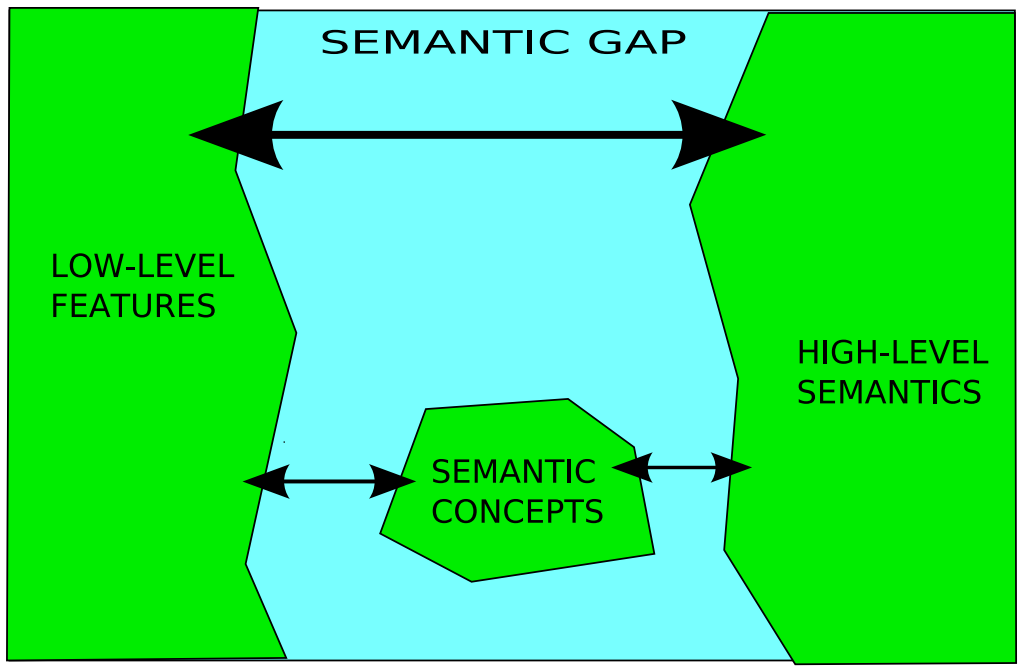
Mats Sjöberg, mats.sjoberg@aalto.fi

Aalto University School of Science
Department of Information and Computer Science

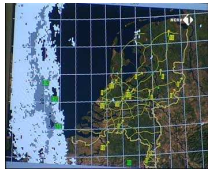
14 June, 2011

Visual concept detection

- A popular approach to alleviate **the semantic gap** is to train a dictionary or ontology of **semantic mid-level concepts**



sports



weather

court

office

meeting

studio

outdoor

building

desert

vegetation

mountain

road

sky

snow

urban

waterscape/ waterfront

crowd

face

person

police/ security

military

prisoner

animal

computer/TV screen

US flag

airplane

car

bus

truck

boat/ship

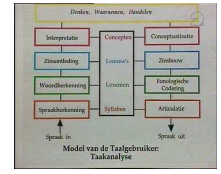
walking/ running

people marching

explosion/ fire

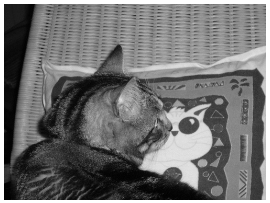
natural disaster

maps



charts

Image category detection: the problem setting



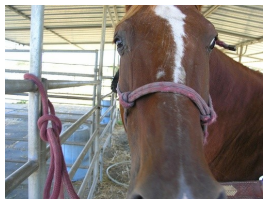
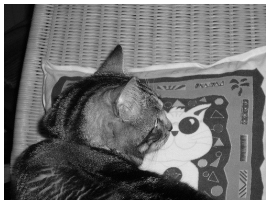
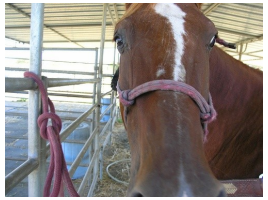


Image category detection: the problem setting







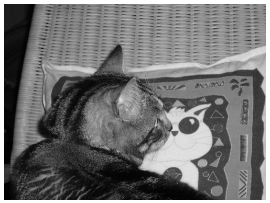
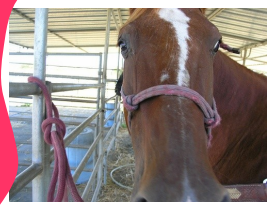


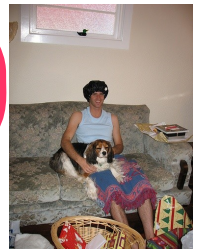
PERSON

Image category detection: the problem setting





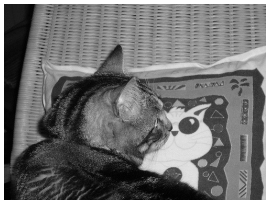




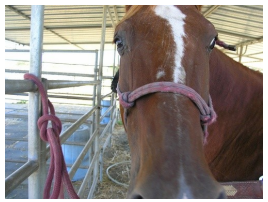


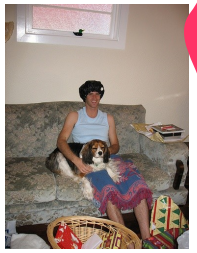


INDOORS

Some categories are more visual than others





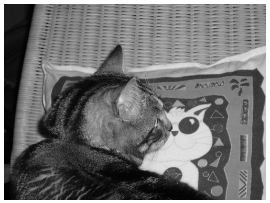
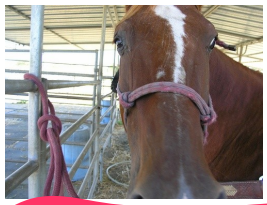


CAR

Some categories are more visual than others







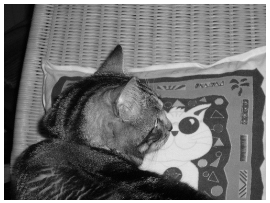
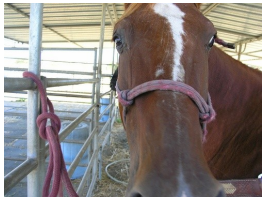


SUMMER

Some categories are more visual than others





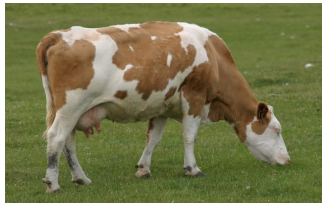




TAKEN IN 2007

Category detection often formulated as a supervised learning problem -> examples

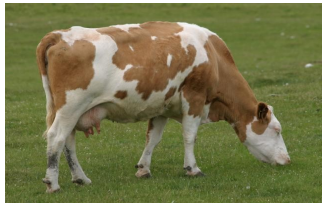
Positive examples



Negative examples

Category detection often formulated as a supervised learning problem -> examples

Positive examples

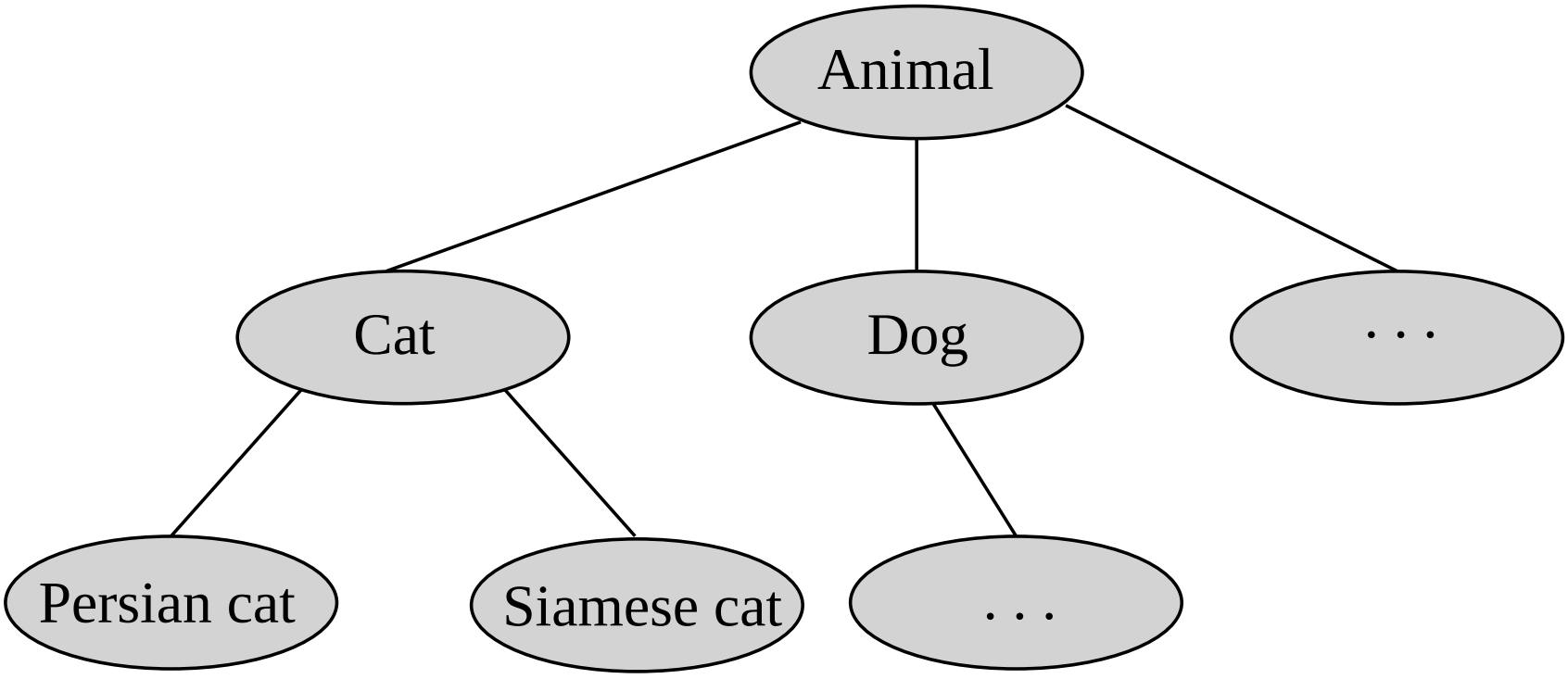


Is this a cow ?

Negative examples

Concept ontologies

- Concepts have structure which can be utilized:



- Also: co-occurrences, context

Semantic feature space

- Given a set of visual concepts $C_1, \ldots, C_K$, we can construct a concept vector for the object x_i:

$$\mathbf{c_i} = \begin{pmatrix} p_{i,1} \\ \vdots \\ p_{i,K} \end{pmatrix},$$

  where $p_{i,j} \in [0, 1]$ is the concept membership score of object x_i in concept C_j
- $\mathbf{c_i}$ can be considered as the **visual context** of x_i