DIRAC

Detection and Identification of Rare Audiovisual Cues

Insperata accidunt magis saepe quam quae speres.
(Things you do not expect happen more often that things you do expect)
Plautus (cca 200 B.C.)

Beyond Novelty Detection: Incongruent Events, when General and Specific Classifiers Disagree

- Daphna Weinshall & Alon Zweig: Hebrew University of Jerusalem, Israel
- Hynek Hermansky: Johns Hopkins University, Baltimore MD, USA.
- Misha Pavel & Holly Jimison: Oregon Health and Science University, Portland OR, USA.
- Luo Jie: Idiap Research Institute/EPFL, Switzerland.
- Frank Ohl: Leibniz Institute for Neurobiology, Magdeburg, Germany.

Novelty Detection

- Machine learning techniques typically attempt to predict the future based on past experience
- An important task is to decide when this is not the right thing to do – the task of novelty detection
- Our contribution:
  - Catalogue different types of novel events based on the combined response of classifiers at different levels of generality
  - Offer a unified framework for incongruent events – events accepted by general classifiers, while being rejected by specific classifiers
  - Identify applications in speech and computer vision, and design appropriate algorithms.

The unexpected is to be expected

- The unexpected (novel) has low posterior probability given past observations, for such reasons as:
  - Poor measurements

  - Low prior probability in a certain context
  - Novel combination of familiar parts

  - Unfamiliar class

Novelty Detection: common practice

A common practice when dealing with novelty is to identify it by **rejection** (or low posterior): declare novelty when no known classifier accepts a test item.

Rejected

=>

Collie is Novel

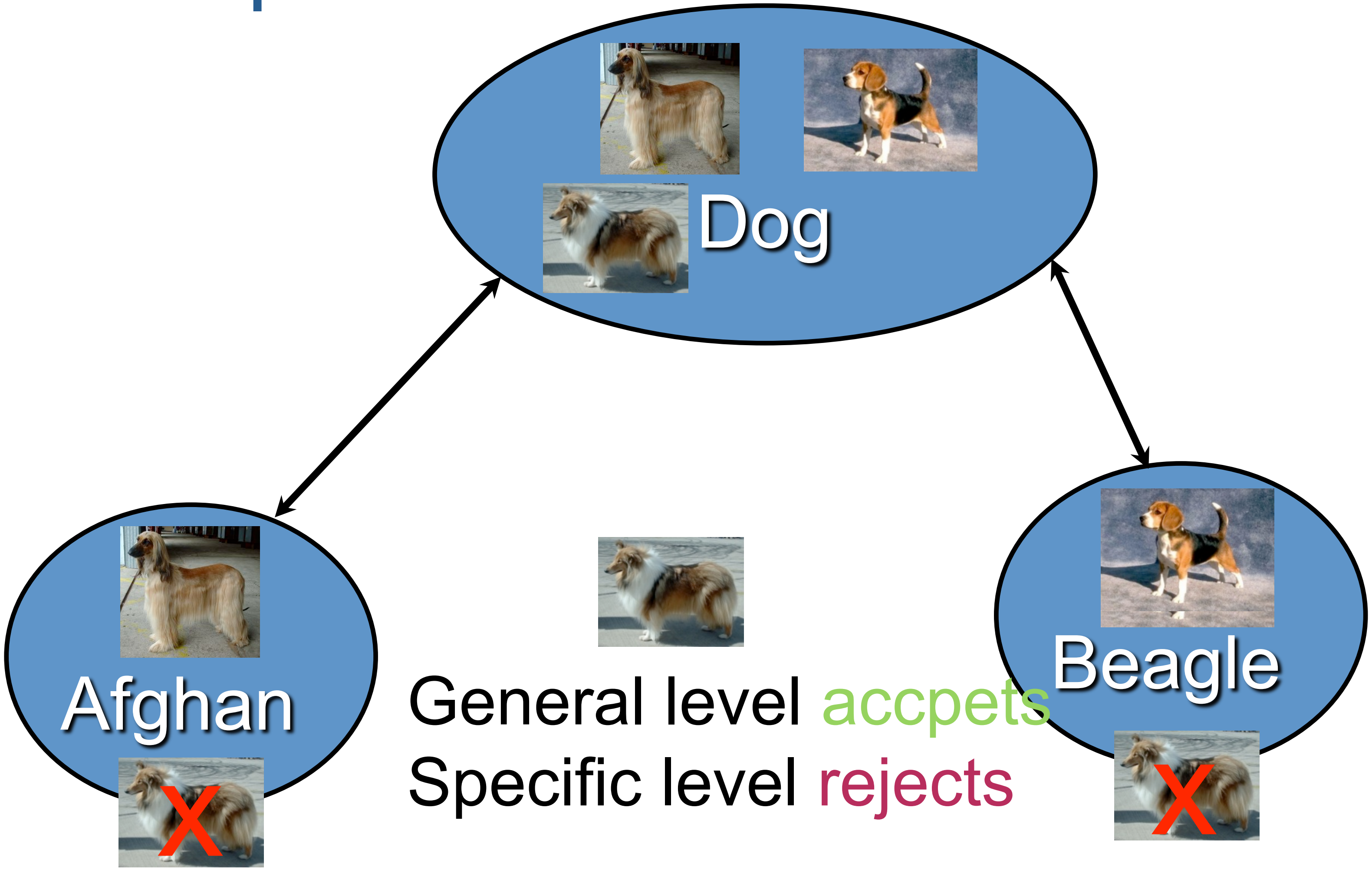
Proposed methods mostly differ in the way known data is modeled and how rejection is achieved.

Incongruent Events

- Novel Incongruent events are detected by the **acceptance** of a general level classifier and the **rejection** of the more specific level classifier.

- Deviation from common practice: we first look for a level of description where the novel event is highly probable.

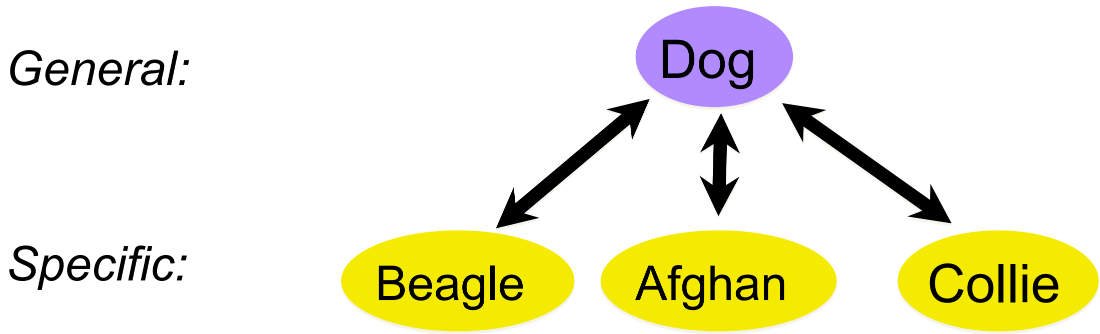
Example:



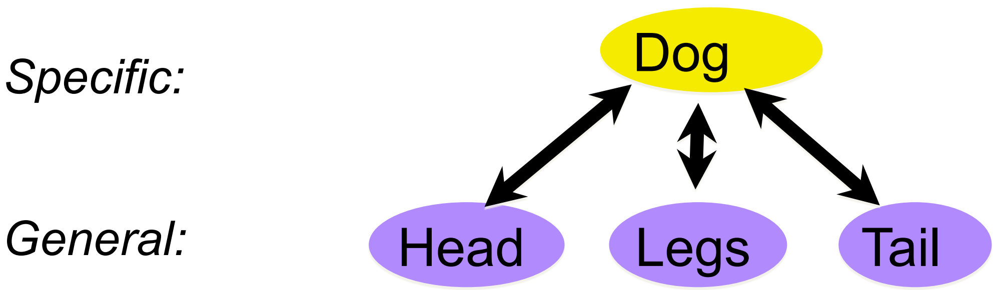
[earlier example: in some face recognition methods, it has been suggested to precede individual face recognition by generic face detection]

Examples of General-Specific relations:

- Class-Membership (as in human categorization) – where objects are categorized at different levels of generality

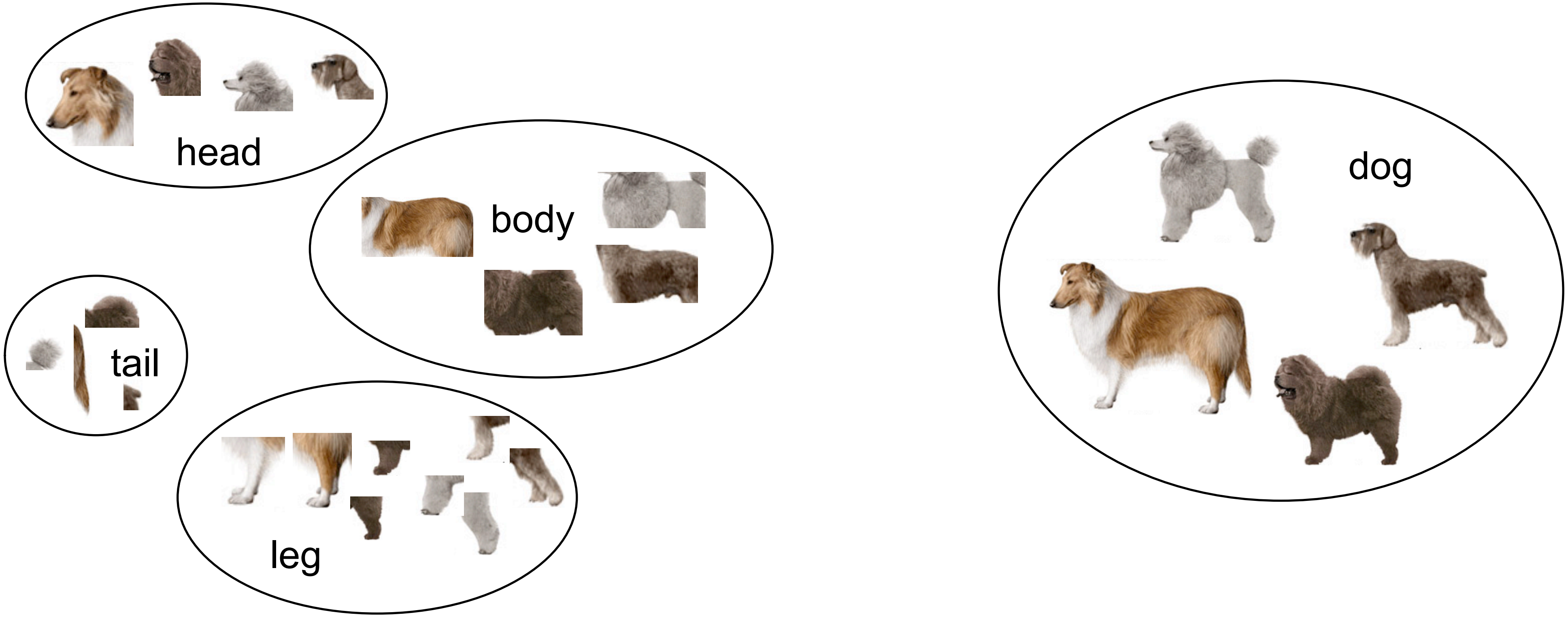


- Part-Whole relationship



[it may seem counter-intuitive that ‘leg’ is more general than ‘dog’; there are more observations showing legs than those of a whole dog]

Part-Whole relationship



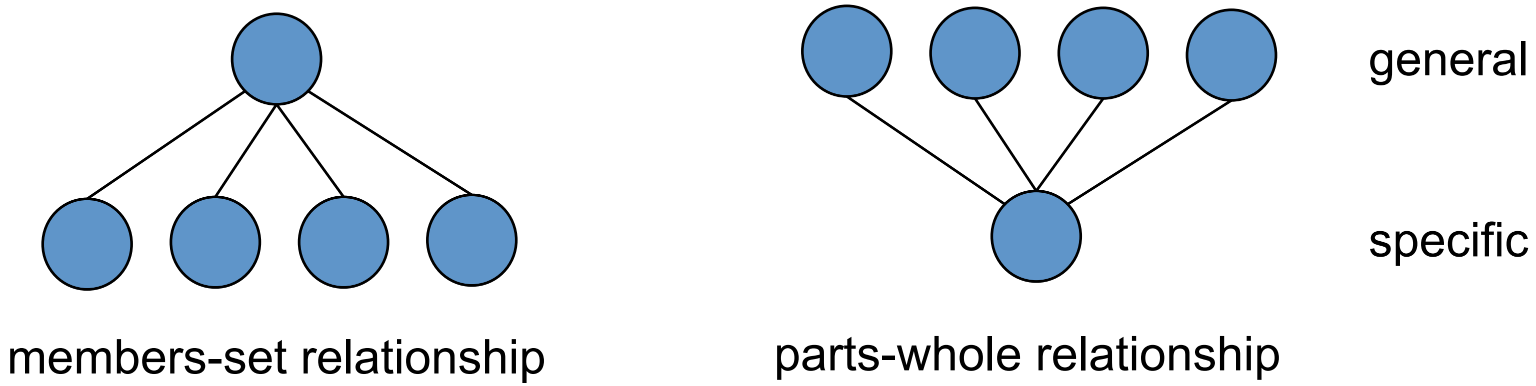
general level

Specific level

dog=body+head+tail+legs

Levels and classifiers:

- There may be one-to-many relations between the general and specific classifiers

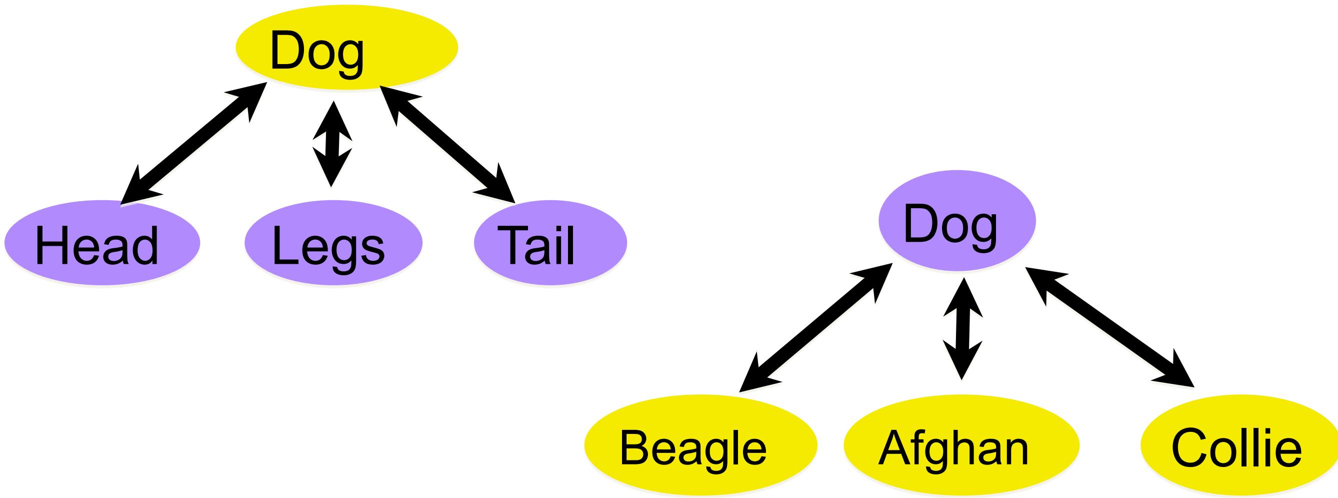


Relations between the levels:

	Specific level	General level	Possible reason
1	reject	reject	noisy measurements, really novel event
2	***reject***	***accept***	***incongruent concept***
~~3~~	~~accept~~	~~reject~~	~~inconsistent , models are wrong~~
4	accept	accept	known concept

Partial Order: A Unified Approach

- The two hierarchies part-whole and class-membership have different one-to-many and many-to-one relations between the general and specific levels.



- In order to deal with both hierarchies in the same framework, we use the notion of partial order.

Partial Order: A Unified Approach

- Concepts are ordered according to the size of the set of events they correspond too: $a \subset b \Rightarrow a \leq b$
- Intuitively speaking, different levels in each hierarchy are related by a partial order: the more specific concept a, which corresponds to a smaller set of events or objects in the world, is always smaller than the more general concept b, which contains all the events in a and more.

Dog = Head∩Legs∩Tail thus Dog ⊂Legs ⇒ Dog ≤ Legs

Dog = Beagle ∪ Afghan ∪ Collie thus Dog⊃Beagle ⇒ Dog ≥Beagle

Partial Order: definitions

Given a class/concept '***a'*** we define:

$\boldsymbol{A^s = \{b \in G, b \preccurlyeq a\}}$ all concepts which are more specific than **'a'**

$\boldsymbol{A^g = \{b \in G, a \preccurlyeq b\}}$ all concepts which are more general than **'a'**

All events which correspond to concept $\boldsymbol{b} \in \boldsymbol{A^s}$ correspond also to concept **a**. *[Each Beagle is also a Dog.]*

All events which correspond to concept **a** correspond also to all concepts $\boldsymbol{b} \in \boldsymbol{A^g}$. *[Each Dog has Legs.]*

Partial Order → multi-level classifiers

For each concept, we construct up to 3 different classifiers:
using the same input data
giving the same (or similar) output on training data

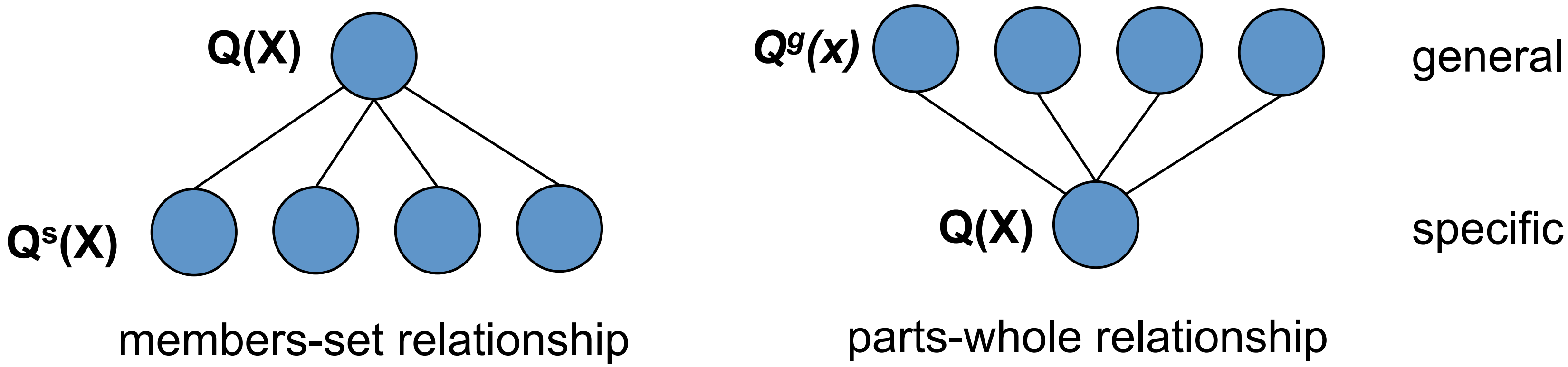
Q(X): sees examples from the concept (implicit knowledge)
Q^s(X)*:* sees examples from the more specific concepts (explicit knowledge)
Q^g(X): sees examples from the more general concepts (explicit knowledge)

Q(x): classifier derived without partial order relations
If $|A^s| > 1$, ***$Q^s(x)$:*** classifier based on the probability of concepts in A^s.
If $|A^g| > 1$, ***$Q^g(x)$:*** classifier based on the probability of concepts in A^g.

<u>We look for disagreement on test data,</u> to find incongruent events:

Observation X *is* incongruent *if there exists a concept for which* **Q^g(X)** accepts and **Q(X)** rejects, or **Q(X)** accepts and **Q^s(X)** rejects

Levels and classifiers:



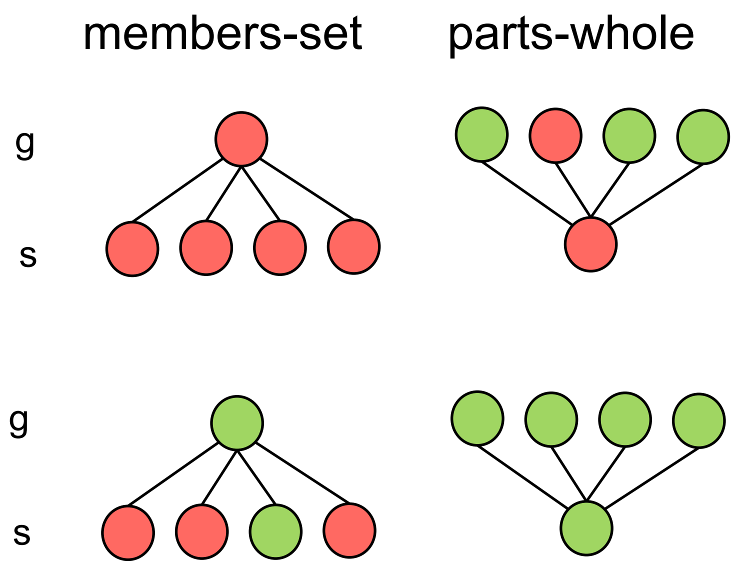
Relations between the levels:

	Specific level	**General level**	**Possible reason**
1	reject	reject	noisy measurements, really novel event
2	***reject***	***accept***	***incongruent concept***
~~3~~	~~accept~~	~~reject~~	~~inconsistent , models are wrong~~
4	accept	accept	known concept

Relations between the classifiers:

- on the training set the two levels have to agree

	Specific level	General level	Possible reason
1	reject	reject	noisy measurements, no concept
4	accept	accept	known concept



one of the general classifiers rejects

all of the specific classifiers reject

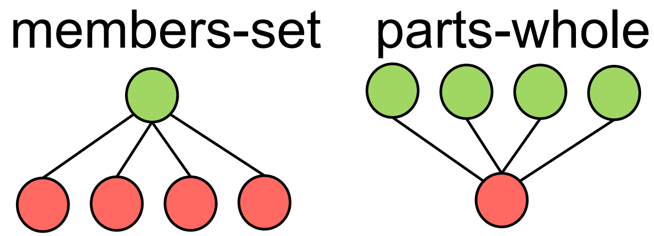
all of the general classifiers accept

one of the specific classifiers accepts

Incongruent Events – the levels disagree on test data

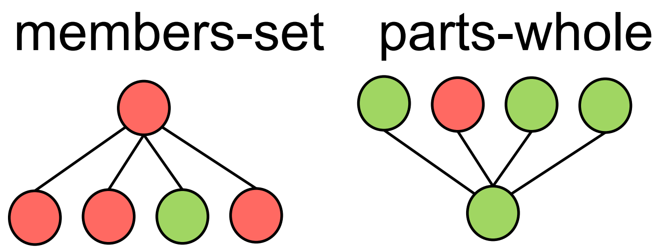
	Specific level	General level	Possible reason
2	***reject***	***accept***	***incongruent concept***

- Item is in the general category, but it doesn't fit any of the sub-categories
- all the parts are there, but the whole isn't there afterall



	Specific level	General level	Possible reason
3	accept	reject	inconsistent with partial order, models are wrong

- Item is not in the general category but one of the members fits one sub-category
- one of the parts is missing, but the whole is still there



Applications

- Unified definition is rather abstract, algorithms are likely to be quite different for the two different hierarchies

- Two different algorithmic implementations

  - Computer vision: **New subclass detection** using Class-membership

  - Speech: **Out Of Vocabulary** word detection using Part-whole membership

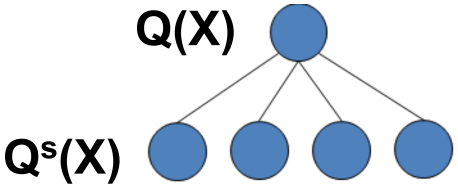
Applications: New subclass detection

Known:

Unknown:

Background:

- Task: Given a sample X, classify it as: a known-subclass, unknown-subclass or background.



members-set relationship

- Two types of classifiers are trained, General classifier: **Q(X),** Specific classifier: **Q^s(X)**.

- An incongruence - acceptance by **Q(X)** and rejection by **Q^s(X),** leads to new subclass detection.

Applications: New subclass detection

Known:

Background:

- **Q(X):**

- The General classifier is trained using the union of the training data from all known subclasses.

- **$Q^s(X)$:**

- Construct a set of discriminative classifiers for all specific subclasses.

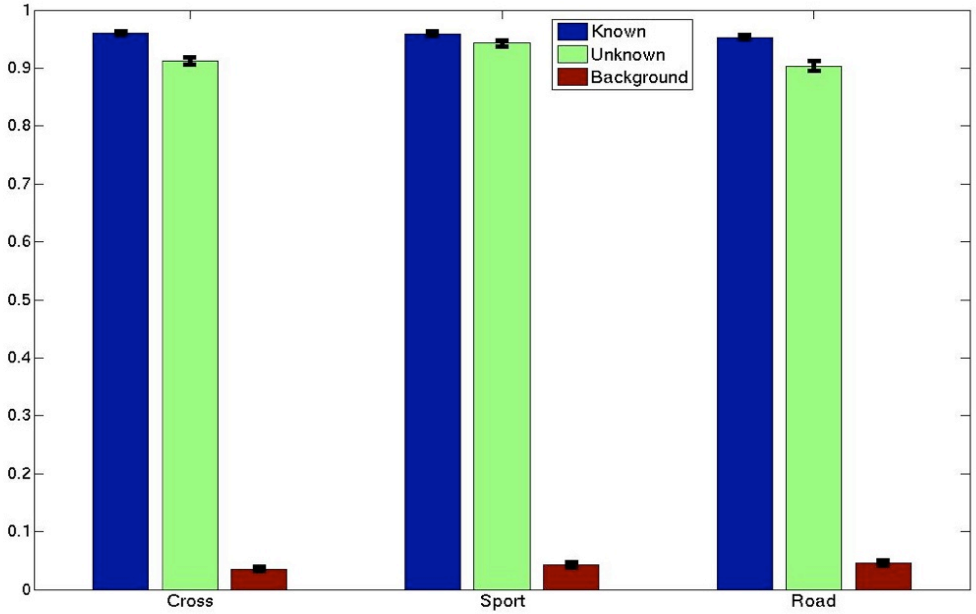
- For each new example: assign the subclass achieving the maximal margin, and return this margin value.

- Compare this margin to a threshold to decide acceptance vs. rejection.

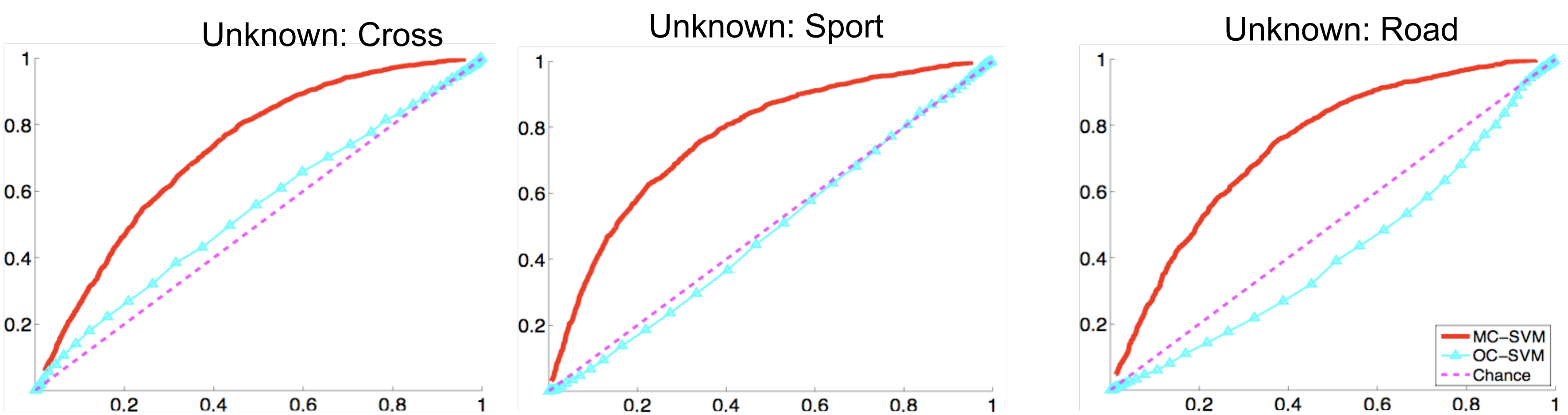
New subclass detection: motorbikes

•Three types of Motorbikes: Cross,Sport & Road. In each set of experiments, one of them is left out as the unknown.

General:



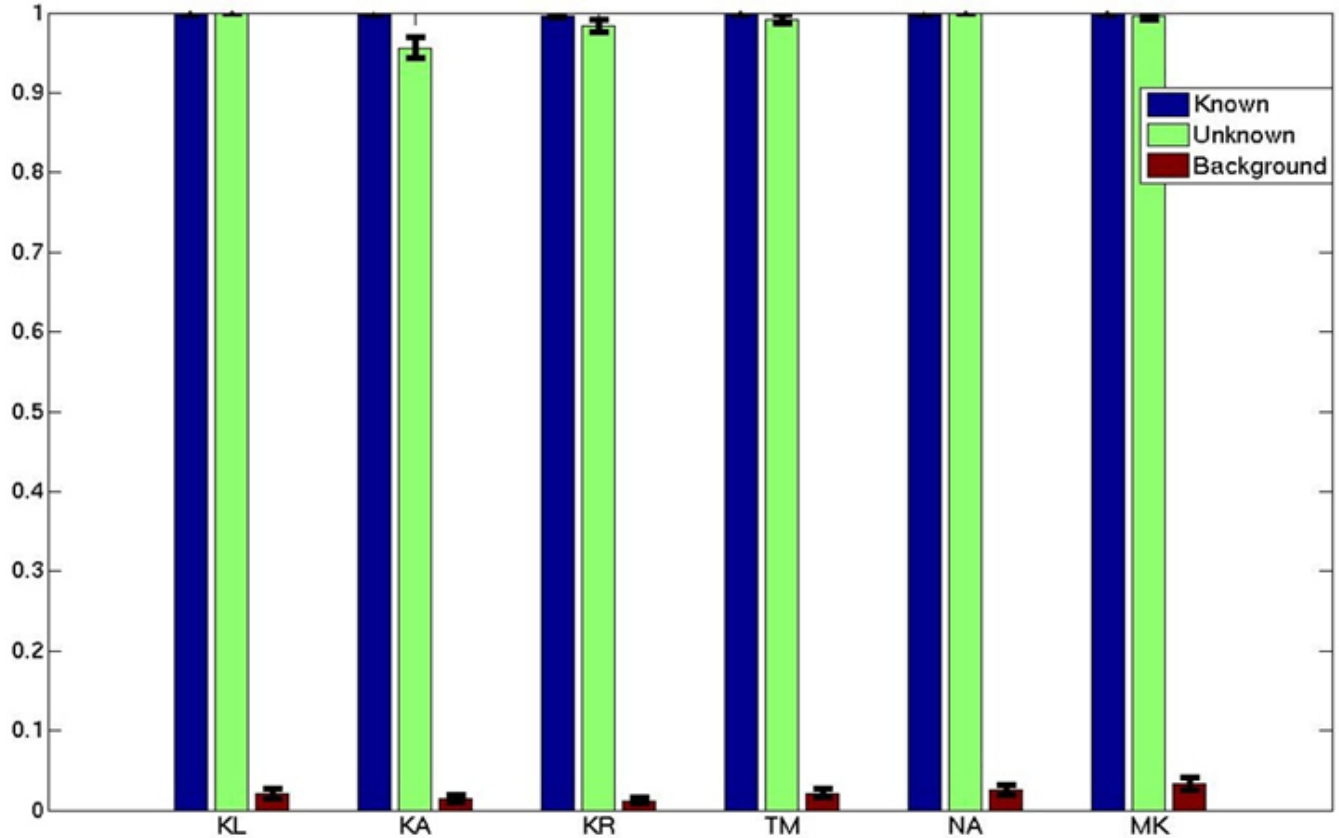
Specific:



New subclass detection: faces

- Six individuals: In each set of experiments, one person is left out as the unknown.

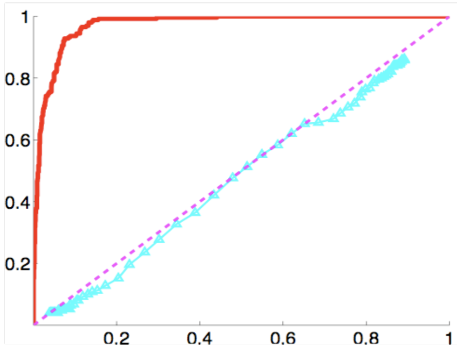
General:



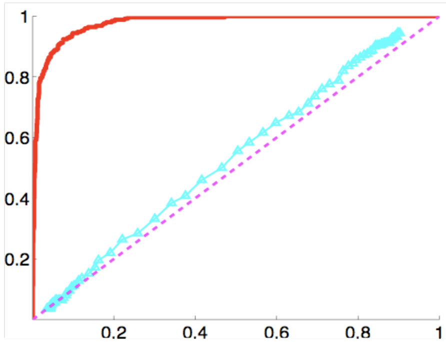
New subclass detection: faces

Specific:

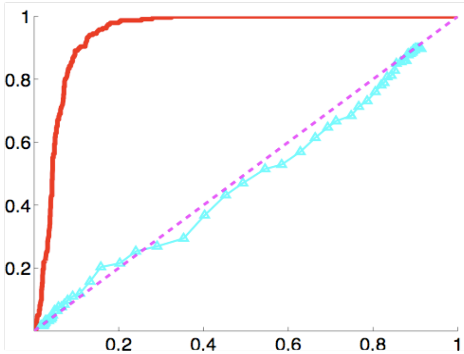
Unknown: KA



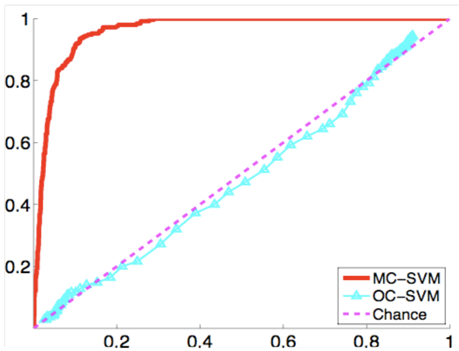
Unknown: KL



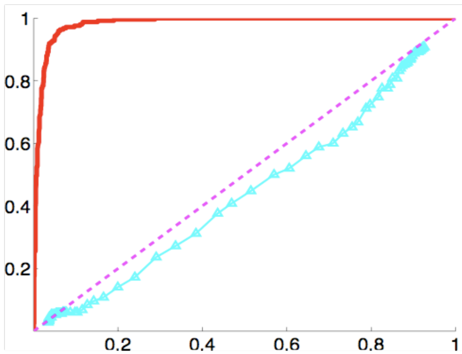
Unknown: KR



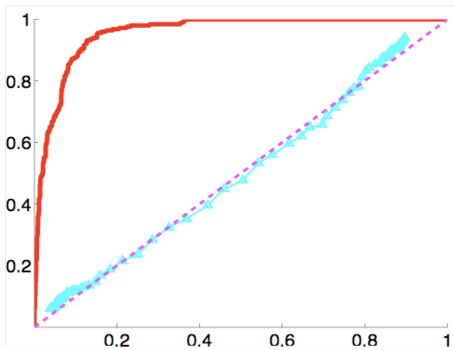
Unknown: MK



Unknown: NA

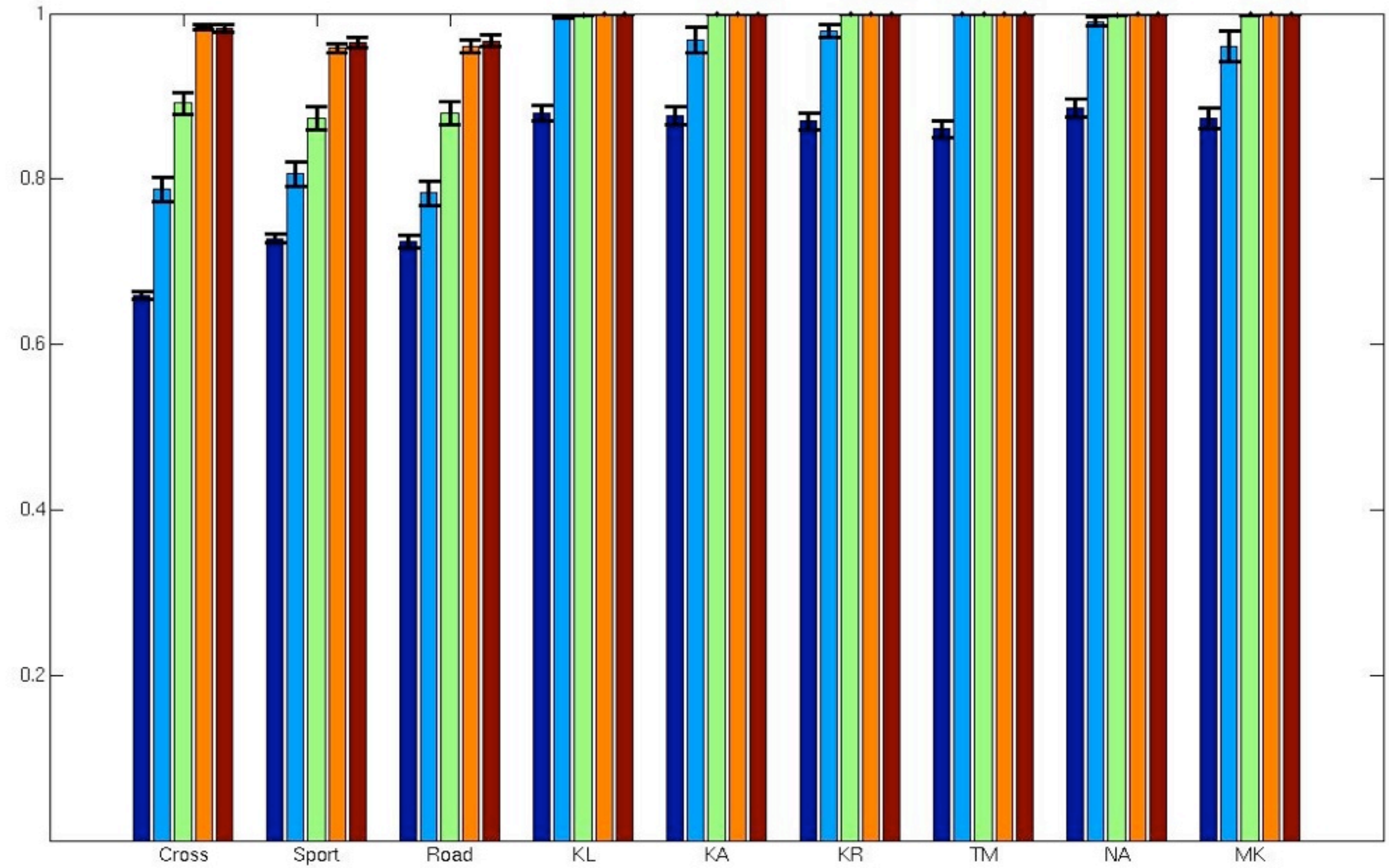


Unknown: TM



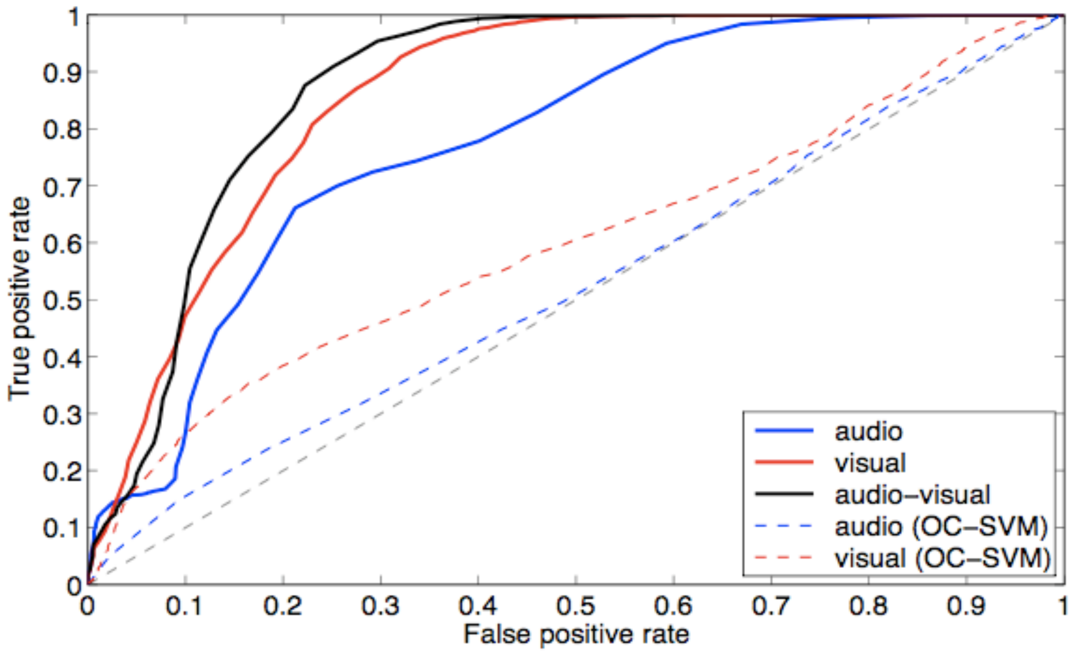
New subclass detection: detection of noisy images

$$\frac{\#\ \text{Specific \& General Reject}}{\#\ \text{Specific Reject}}$$



New subclass detection: Speaker verification

- Six known and 11 unknown individuals, photographed while approaching the camera and speaking to a microphone.
- General level Face and Speech classifiers.
- Specific individual classifiers.
- Fusion was done by using a threshold over the normalized average margin of both modalities.

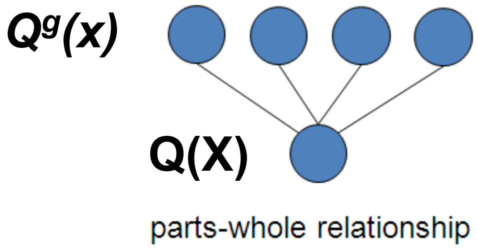


Out Of Vocabulary word detection

- An Out-Of-Vocabulary word is a word that doesn't appear in the dictionary.

- Motivation: the appearance of such a word in an utterance typically carries more information than the rest of the words in the utterance.

- This is a part-whole example – utterances are combinations of words.



- Two ways for computing the probability of an utterance:

-General level: using generic Phoneme recognizers

-Specific level: using Constrained language model

Out Of Vocabulary word detection

-General level: using generic Phoneme classifiers

$$Q^g(X) = p(X) = \sum_u p(X|u)p(u) \geq p(X|\bar{u})p(\bar{u}) = p(X|\bar{u}) \prod_k p(w_k)$$

-Specific level: Constrained language model

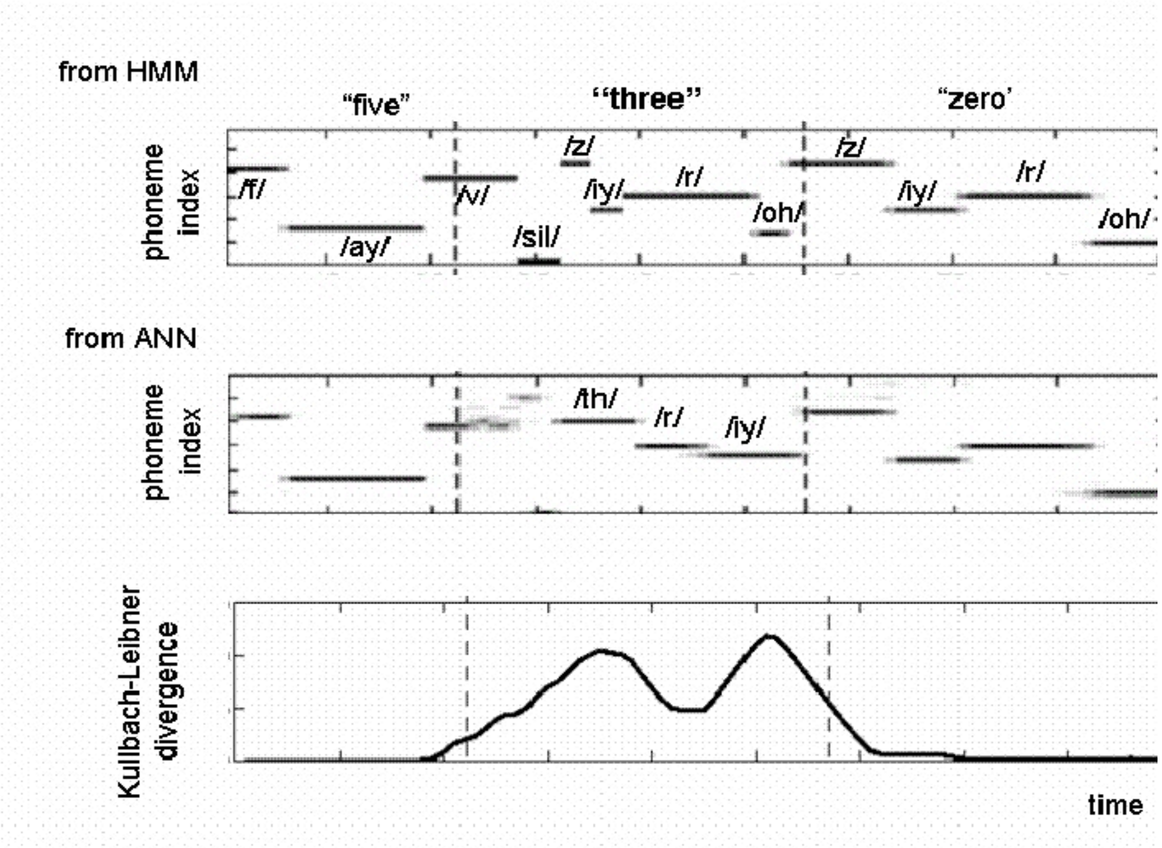
$$Q(X) = p(X|L) = \sum_u p(X|u, L)p(u|L) \approx p(X|\bar{u}, L)p(\bar{u}|L) = p(X|\bar{u}) \prod_k p(w_k|L)$$

Incongruency detection algorithm: compute the KL-divergence between the probability distributions over phonemes (posteriograms) for each word, induced by **Q(X)** and **Q^g(X)**. Identify incongruence when this distance is unusually large.

Problem: divergence may fail when models are wrong, or when the two posteriograms differ simply because the classifiers implicitly reject by predicting a different outcome.

Experiment 1

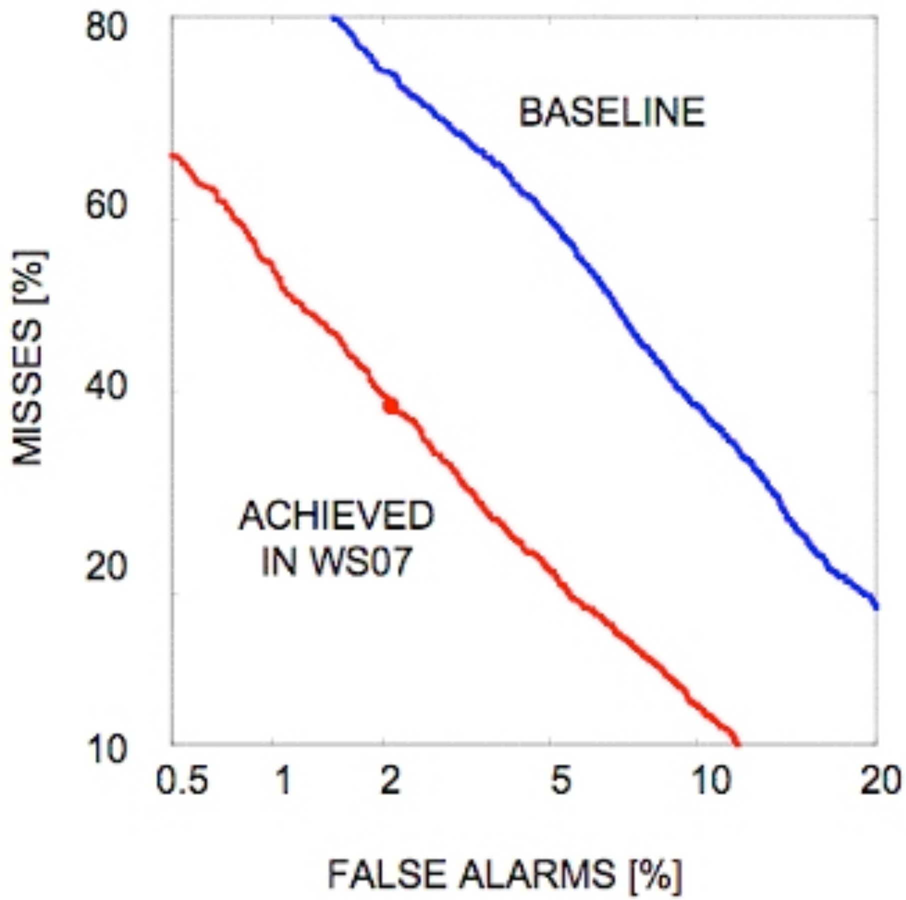
- An HMM constrained recognizer based on a lexicon without the word ‘three’
- An unconstrained (no lexicon) phoneme based recognizer.
- The constrained recognizer forced the recognition of ‘three’ as ‘zero’
- Posterior probabilities of phonemes (posteriograms):



- Big divergence on the OOV word!!

Out Of Vocabulary detection

- Test on Wall Street Journal data set
- 20% least frequent words left out as OOV
- Compared to state-of-the-art Cmax technique



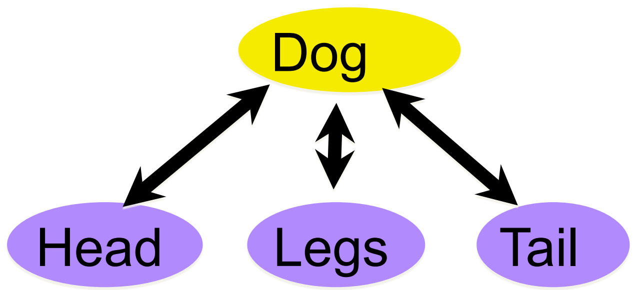
Summary

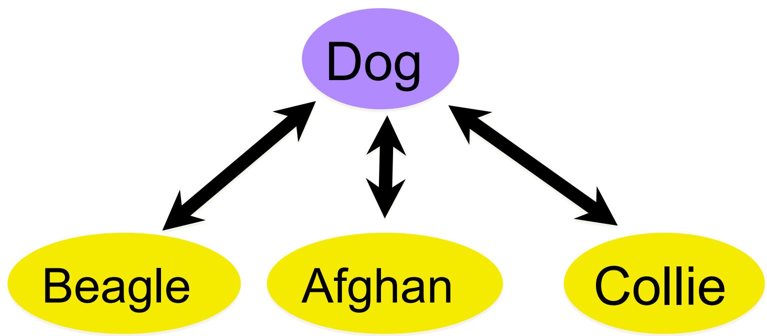
- Observing that there are different types of novel/unexpected events, we focus on a specific type of novel events, namely: **Incongruent events**.
- Contrary to common practice we first look for a level of description where the novel event is highly probable.
- Novel **Incongruent events** are detected by the **acceptance** of a general level classifier and the **rejection** of the more specific level classifier.
- We present a general framework based on the notion of partial order on labels for the detection of such novel events. Different types of labeled hierarchies such as **part-whole** and **class-membership** are captured by this framework.
- We demonstrate two different algorithmic implementations of this framework for two different types of novelty: new subclass detection (vision), and Out Of Vocabulary word detection (speech).

Thanks

Partial Order: A Unified Approach

- The two hierarchies part-whole and class-membership induce constraints on the observed features in different ways.
- In the class-membership hierarchy, a parent class admits higher number of combinations of features than any of its children, i.e., the parent category is less constrained than its children classes.
- In contrast, a parent node in the part-whole hierarchy imposes stricter constraints on the observed features than a child node.
- Our contribution: we deal with both hierarchies using the same framework





Partial Order: implied recognizers

We construct different recognizers
using the same input data
giving the same (or similar) output on training data
Using either
Q(x): only implicit knowledge extracted from training data
$Q^s(x)$, $Q^g(x)$: explicit knowledge via the partial order

Q(x): a classifier for class '**a**', derived from training data without using the partial order relations.

If $|A^s| > 1$, ***$Q^s(x)$:*** a classifier for class '**a**' which is based on the probability of concepts in A^s.

If $|A^g| > 1$, ***$Q^g(x)$:*** a classifier of class '**a**' which is based on the probability of concepts in A^g.

<u>We look for disagreement on test data,</u> to find incongruent events:
Observation X *is* incongruent *if there exists a concept* '**a**' *such that*
$Q^g(X) \gg Q(X)$ or $Q(X) \gg Q^s(X)$.