Concepts and Categories
Informatics 1 CG: Lecture 16
Chris Lucas

Note:
We're focusing on concepts that are mental representations of classes of objects or events.

You might have concepts of “skiing”, “justice” or “wanderlust”, but the category “things we’re talking about today” does not include them.

Today
(1) Revisit theories of categorisation and
(2) Connections to inductive bias and generalisation

The uses of categorisation
What are categories good for?
1. Efficient representation
2. Communication
3. Generalisation

Theories of categorisation
- Definitional (or “classical”) theory
- Similarity-based approaches
  - Prototype theory
  - Exemplar theory
- Theory theory

Definitional (or “classical”) theory
Categories have necessary and sufficient features, e.g.,
“bachelor” ⇔ unmarried & adult & male
Definitional (or “classical”) theory

Pros:
• Intuitive; economical; easy to communicate.

Cons:
• Good definitions are hard to find
  • Is the pope a bachelor?
  • What about an unmarried person in a single-partner long-term relationship?
  • What’s “male”? “Adult”? 
• Can’t explain typicality effects or fuzzy boundaries
• Where do definitions come from?

Similarity-based theories: **Prototype theory**

Membership is based on similarity to a category prototype – a summary representation, usually taken to be an average.

**Pros:**
• Economical representation.

**Cons:**
• Has trouble capturing complex category structure.

Similarity-based theories: **Exemplar theory**

Membership is based on similarity to known category members.

**Pros:**
• Flexible representation; can represent categories that don’t have a single mode and complex category boundaries

**Cons:**
• Not economical

Similarity-based theories: **Hybrids**

Can we combine the advantages of prototype and exemplar theories?

**Idea:** lexical concepts can correspond to many clusters of entities, e.g.:
• fluffy white cats,
• tabby cats,
• that one green cat
Similarity-based theories: **Hybrids**

We can express both exemplar and prototype theories this way, and treat categorisation as a problem of density estimation. Can we use this to understand human categories and concepts in detail?

Similarity-based theories

**Challenges to all similarity-based theories:**

- Where do features come from?

Similarity-based theories

**Challenges to all similarity-based theories:**

- Compositionality

Similarity-based theories

- Is Sweden, Poland, or Hungary most similar to Austria?
  - Sweden (49%) > Hungary (36%)

- Is Sweden, Norway, or Hungary most similar to Austria?
  - Hungary (60%) > Sweden (14%) [irrelevance]

Similarity-based theories

- Discourse context

  - Within-individual variability
Similarity-based theories

- Variance effects:

Theory theory

- Category membership depends on causal and explanatory features.
  - Causal features are more important than surface features, e.g.,
    - Function > appearance (for adults, at least)
    - Cat DNA > Cat-like appearance
  - Does everything have one natural category?
  - Can we think of category labels as features?

Categories and generalisation

A typical generalisation problem involves:

- A new case and some data about it,
- Previously-observed cases,
- Some background and contextual information,

We want to draw conclusions about the new case.

Categories and generalisation

We might want to know different things:

- Is it edible?
- Will it try to eat us?
- What's its display resolution?
- How should we label it?

Category-based induction

Example:

if pelicans have a choroid membrane in their eyes and
albatrosses have a choroid membrane in their eyes, so
all birds have a choroid membrane in their eyes?

(Osher et al., 1990)

Category-based induction

Phenomena:

- Premise typicality
- Premise diversity
- Conclusion specificity
- Premise monotonicity*
- Inclusion fallacy

(Osher et al., 1990)
Category-based induction

Phenomena:
• Premise typicality

Robins have feature
Birds have feature

→

Penguins have feature
Birds have feature

(Ohlsson et al., 1990)

Category-based induction

Phenomena:
• Premise diversity

Hippos have feature
Hippos have feature

→

Mammals have feature
Mammals have feature

(Ohlsson et al., 1990)

Category-based induction

Phenomena:
• Conclusion specificity

Bluejays have feature
Falcons have feature

→

Bluejays have feature
Falcons have feature

→

Birds have feature
Animals have feature

(Ohlsson et al., 1990)

Category-based induction

Phenomena:
• Premise monotonicity*

Hawks have feature
Sparrows have feature

→

Eagles have feature
Eagles have feature

→

Birds have feature
Birds have feature

(Ohlsson et al., 1990)

Category-based induction

Phenomena:
• Premise monotonicity*

Sparrows have feature
Eagles have feature

→

Robins have feature
Robins have feature

→

Ostriches have feature
Ostriches have feature

(Ohlsson et al., 1990)

Category-based induction

Phenomena:
• Inclusion fallacy

Sparrows have feature
Eagles have feature

→

Robins have feature
Robins have feature

→

Birds have feature
Birds have feature

(Ohlsson et al., 1990)
Theories

Similarity-based accounts of category-based induction:
• Tversky’s contrast model (1977):
  Feature overlap determines salient features.
• Osherson et al. (1990):
  Weighted combination of similarity and coverage
  Assumes stable, hierarchical categories
• Connectionist (neural network) model (Sloman, 1993):
  Proportion of shared features between premises and conclusion
  Estimated with neural network
  [and others]

Theory theory strikes again!

Causal knowledge drives category-based induction.

Examples:
• If X eats Y, they’re more likely to share disease.
• If X is taxonomically related to Y, they’re more likely to share bone types.
• If X is the same weight as Y, they’re likely to need similar amounts of sodium in their diet.

How can we use and combine these kinds of knowledge?

For one proposal, see [1].


Summary

Similarity is at the heart of prototype and exemplar theories, but it’s a complex concept in its own right.
• Context matters! (What’s being compared, goals, ...)
• Category variability
• Trade-off between expressiveness and economy; hybrid models can help

Categories help us generalise
• Category-based induction:
  features of some categories or exemplars → inferences about others

“Theory theory” issues and questions remain