Acta Psychologica 133 (2010) 216–233

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Category and feature identification

Charles Kemp [a,*], Kai-min K. Chang [b], Luigi Lombardi [c]

[a] Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[b] Language Technologies Institute, Carnegie Mellon University, Pittsburgh PA 15213, USA
[c] Dipartimento di Scienze della Cognizione e della Formazione, Università di Trento, I-38068 Rovereto, Italy

ARTICLE INFO

Article history:
Received 17 December 2008
Received in revised form 7 November 2009
Accepted 16 November 2009
Available online 18 January 2010

PsycINFO classification:
2340
2343
4100

Keywords:
Inductive reasoning
Identification
Bayesian models
Information integration
Semantic cognition
Naming to description

ABSTRACT

This paper considers a family of inductive problems where reasoners must identify familiar categories or features on the basis of limited information. Problems of this kind are encountered, for example, when word learners acquire novel labels for pre-existing concepts. We develop a probabilistic model of identification and evaluate it in three experiments. Our first two experiments explore problems where a single category or feature must be identified, and our third experiment explores cases where participants must combine several pieces of information in order to simultaneously identify a category and a feature. Humans readily solve all of these problems, and we show that our model accounts for human inferences better than several alternative approaches.

Suppose that you are watching a German nature program and that you pick up enough of the narrative to learn that a *Schmetterling* is colorful, has wings, and has antennae. Can you guess what a *Schmetterling* might be? Similarly, suppose that you learn that zebras and tigers are both *gestreifet*. Can you guess what *gestreifet* might mean? We will refer to both of these problems as *identification* problems. In the first case, you need to identify a category – namely, *butterfly*. In the second case, you need to identify a feature – namely, *striped*. Problems like these draw on semantic knowledge about animals and their features, and this paper will consider how this knowledge can be used to address identification problems.

As our opening examples suggest, category identification and feature identification are problems regularly faced by second-language learners. In many cases these learners will already have concepts like *butterfly* and *striped*, and their task is to map novel labels onto these concepts. Identification, however, may play an equally critical role in first-language acquisition. Before learning her first few words, a child may already have formed a category that includes creatures like the furry pet kept by her parents, and learning the word "cat" may be a matter of attaching a new label to this pre-existing category (Chomsky, 1991; Fodor, 1975; Mervis, 1987). Bloom (2000) summarizes this proposal by suggesting that "much of what goes on in word learning is establishing a correspondence between the symbols of a natural language and concepts that exist prior to, and independently of the acquisition of that language" (p. 242).

This paper develops a probabilistic framework that can address a broad family of identification problems. Like all inductive problems, identification problems can only be solved if a learner relies on background knowledge, and our approach offers a formal characterization of the knowledge that guides category and feature identification. We propose that this knowledge is stored in a semantic repository that includes information about the relationship between categories and features (for instance, butterflies have wings) along with information about the frequency with which different categories and features are encountered (a random speaker is more likely to refer to dogs or cats than to chameleons or llamas). We make these ideas concrete by describing a repository built from the Leuven Natural Concept Database (De Deyne et al., 2008).

Prior knowledge plays a critical role in inductive reasoning, but this knowledge must be combined with evidence in order to solve inductive problems. Often multiple pieces of evidence are available,

* Corresponding author. Address: Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Baker Hall 340T, Pittsburgh, PA 15213, USA. Tel.: +1 412 268 4680.
E-mail address: ckemp@cmu.edu (C. Kemp).

0001-6918/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2009.11.012

and a reasoner must integrate all of this information. Several accounts of information integration can be found in the psychological literature (Anderson, 1981), and different approaches combine multiple pieces of evidence by adding (Lombardi & Sartori, 2007), multiplying (Medin & Schaffer, 1978; Oden & Massaro, 1978) or taking the maximum (Osherson, Smith, Wilkie, Lopez, & Shafir, 1990) of a set of numerical scores. We will argue that probabilistic inference provides a principled account of information integration that avoids arbitrary choices of functions like sums and products.

The inductive problems we consider and the modeling approach we pursue both build on previous contributions to the psychological literature. The problem of identification is related to the work of Lombardi and Sartori (2007) (see also Sartori & Lombardi, 2004) who developed a computational account of category identification that is known as the additive relevance model. These authors report that their model performs better than a simple Bayesian alternative, but their analysis was based on sparse feature matrices that may not adequately capture what people actually know about categories and their features. Our results suggest that a Bayesian account of category identification performs better than the additive relevance approach when both are supplied with a semantic repository that better captures the knowledge that people bring to the problem.

Several psychologists have developed probabilistic models of inductive reasoning (Anderson, 1990; Heit, 1998; Shepard, 1987) and our approach continues within this general tradition. Of the many probabilistic models that have been developed, our approach is related most closely to models of categorization (Anderson, 1991) and generalization (Kemp & Tenenbaum, 2009) that attempt to explain how inferences about novel objects and properties are guided by semantic knowledge. The identification problems we consider are somewhat different, but our approach is consistent with the idea that probabilistic inference is a domain-general principle that helps to explain how humans solve many inductive problems. Although we focus throughout on identification, we return to the relationship between identification and other inductive problems in the general discussion.

1. A probabilistic account of category and feature identification

This paper will focus on the three identification problems in Table 1. Each problem consists of a list of statements about animal categories and their features, and each list includes a hidden category C, a hidden feature F or a hidden category and a hidden feature. In each case the task of the reasoner is to identify the hidden items. Although the problems in Table 1 are simple enough to be experimentally tractable, they are inspired in part by the real-world inductive challenge faced by first- and second-language learners. In real-world identification problems, the hidden category or feature will typically be introduced as an unfamiliar component of a linguistic utterance (e.g. *Punda milia* have stripes), and the task of the learner is to identify the meaning of this novel word or phrase.

This paper will develop a unified probabilistic model that addresses all three of the problems in Table 1. For each of these problems, our model specifies a probability distribution over the values of the hidden items given the items that have been observed. We propose that humans choose categories and features that have high probability according to these distributions.

Table 1
Three identification problems. Each problem asks a reasoner to identify a category C, a feature F, or a category and a feature.

Problem	Form	Example	Example response
Category identification	Cs have $\{f_1, \ldots, f_n\}$	Cs have stripes	C = *zebra*
Feature identification	$\{c_1, \ldots, c_m\}$ have F	Rabbits have F	F = *long ears*
Joint identification	$\{c_1, \ldots, c_m\}$ have F	Rabbits have F	F = *fur*
	Cs have F	Cs have F	C = *tiger*
	Cs have $\{f_1, \ldots, f_n\}$	Cs have stripes	

To formally specify these distributions we take a generative approach. More precisely, we specify a probabilistic procedure for generating identification problems like the examples in Table 1. Suppose that we start with a semantic repository that captures knowledge about animal categories and their features. We will specify a procedure that samples a list of statements from this repository, including, for example, the statement that "zebras have stripes". We now assume that some of the categories and features in the sampled statements are hidden – for example, "zebras have stripes" might become "Cs have stripes". Given this procedure for generating identification problems, we can now use Bayesian inference to work backwards and identify the hidden items in any given problem.

The semantic repository plays a critical role in this approach and must specify two kinds of distributions. First, it must specify a prior distribution $p(c)$ over categories and a prior distribution $p(f)$ over features. These distributions can capture factors like the familiarity of a category and the frequency with which a feature is thought about. For example, in most contexts a familiar category like *dog* should receive higher prior probability than a category like *chameleon*. The semantic repository must also provide two additional distributions: $p(f|c)$ which specifies the probability that a person will choose *f* when asked to list a feature of category *c*, and $p(c|f)$ which specifies the probability that a person will chose *c* when asked to list an animal category that has feature *f*. For example, $p(barks|dog)$ should be greater than $p(breathes\ air|dog)$, since *barks* is the more characteristic feature of dogs, and $p(breathes\ air|dog)$ should be greater than $p(has\ wings|dog)$, since dogs breathe air but do not have wings. Similarly, $p(robin|has\ wings)$ should be greater than $p(penguin|has\ wings)$, which in turn should be greater than $p(dog|has\ wings)$.

The four distributions $p(c)$, $p(f)$, $p(c|f)$ and $p(f|c)$ can be used to generate many kinds of identification problems. Here we focus on three problems that we refer to as category identification, feature identification, and joint category and feature identification.

1.1. Category identification

The first problem in Table 1 requires a reasoner to identify an animal category given one or more features of the category. For example, the reasoner might be informed that "Cs have stripes and hooves" and asked to identify Category C. As shown in Fig. 1a, we assume that problems of this kind are generated by sampling a category c (here $c = zebra$) from the prior distribution $p(c)$, then sampling n features from the distribution $p(f|c)$. As a final step, the value of c is hidden and the reasoner is asked to identify this category.

We model this inference using the posterior distribution $p(c|f_1, \ldots, f_n)$, or the distribution over categories given the features that have been observed. This distribution can be written as

$$p(c|f_1, \ldots, f_n) \propto p(f_1, \ldots, f_n|c)p(c) \tag{1}$$

$$= \prod_{j=1}^{n} p(f_j|c)p(c), \tag{2}$$

where the right hand side is expressed using distributions specified by the semantic repository ($p(f|c)$ and $p(c)$). Eq. (2) combines two criteria: the hidden category should have high prior probability ($p(c)$ should be high), and should also be consistent with the observed features ($p(f_j|c)$ should be high for each observed feature j). Note that Eq. (2) follows from Eq. (1) only if the features f_1 through f_n are conditionally independent given the hidden category c. We