A Model of Concept Generalization and Feature Representation in Hierarchies

Timothy N. Rubin (trubin@uci.edu)
Department of Cognitive Sciences, University of California, Irvine

Matthew D. Zeigenfuse (mzeigenf@msu.edu)
Department of Psychology, Michigan State University

Mark Steyvers (msteyver@uci.edu)
Department of Cognitive Sciences, University of California, Irvine

Abstract

We present a novel modeling framework for representing category exemplars and features. This approach treats each category exemplar as a probability distribution over a hierarchically structured graph. The model jointly learns the mixture of each exemplar across categories in the graph, and a feature representation for each node in the graph, including nodes for which no data is directly observed. We apply this model to two distinct types of data: (1) Animal by Feature matrices from the Leuven Natural Concept Database, and (2) documents from Wikipedia. We demonstrate that this model is useful for learning feature representations for nodes in the graph that are not assigned any data (i.e. for generalization to new categories). Additionally this model improves the specificity of feature representations for the nodes with observed data by explaining away more general features to parent nodes within the graph. Furthermore, we illustrate that this model is useful for understanding additional psychological aspects of concept representation, such as typicality ratings.

Keywords: Concepts; Category Learning; Graphical Models; Hierarchical Models; Bayesian; Generalization

Suppose you were presented with an unfamiliar concept, *blorg*. Despite having no experience with *blorg*s, if you were told a *blorg* is an animal you would know that it eats and that it breathes. If you were told a *blorg* is a mammal you would know that it has hair and births live young, as well as that it eats and breathes, since mammals are also animals. Finally, if you were told that a *blorg* is a dog you would know that it can bark, as well as that it possesses all of the features of animals and mammals. Clearly, both the categories to which an exemplar belongs and the hierarchy in which those categories reside carry considerable information about the features of that exemplar. Alhough some work has been done looking at how people associate features to a particular category (e.g., Kemp & Tenenbaum, 2009; Austerweil & Griffiths, 2009; Zeigenfuse, 2010), it is unclear how people learn to associate features to levels in a category hierarchy, particularly when they must generalize to categories in the hierarchy for which there is no observed data.

In this paper, we present a rational model of how people jointly learn to associate features with a particular level within a hierarchy, and to learn distributed representations of exemplars across this hierarchy. This model begins with feature representations of exemplars of categories, and the category structure of the domain to which the exemplars belong. It learns the features associated with each category within the structure (even for categories for which there are no exemplars), as well as a distributed representation for each exemplars across multiple levels of abstraction within the hierarchy. This approach differs fundamentally from many other approaches to modeling hierarchical relationships between categories in that learns a *distributed representation* for each exemplar for a category. Specifically, an underlying assumption behind many such approaches is that an exemplar's features are inherently tied to only the category to which the exemplar belongs. This assumption underlies many classical approaches to modeling hierarchical relationships, such as hierarchical clustering methods (e.g., Shepard, 1980), as well more recent advances which learn the basic structural *form* of these relationships (of which a hierarchy is just one possibility), in addition to the graph itself (Kemp & Tenenbaum, 2008). One notable exception is the approach to distributed representations of semantic memory proposed by Collins & Quillian (1969). Although Collins & Quillian (1969) did not address the problem of learning their proposed representations, and the topic of their paper is somewhat different from our own, the underlying approach of the model we present is very much in the spirit of their work.

We apply our model to two highly distinct datasets. First, using Animal by Feature matrices from the Leuven Natural Concept Database (de Deyne et al., 2008), we learn featural representations of animals at the species (e.g., DOGS), animal-category (e.g., MAMMALS), and domain (i.e., ANIMALS) levels of abstraction (despite there being no data directly assigned at either the animal-category or domain level). We then show that the representation of exemplars as probability distributions across the hierarchy naturally captures psychological phenomena such as an animal's perceived "typicality" for a category, which has been shown to be a fundamental property of category representation (e.g., Rosch & Mervis, 1975). We additionally apply our model to documents from a subset of the Wikipedia category structure, in order to demonstrate that our approach is applicable to noisy, real-world data, represented within a more convoluted hierarchical structure that spans multiple domains with a wide range of category specificity.

A Mixture Model for Representing Exemplar Features over Graph Hierarchies

In this section, we present a model for learning feature representations for categories using a framework related to the Topic Model (Blei et al., 2003). The Topic Model was originally presented as an unsupervised learning method for finding low-dimensional representations of text corpora. In psychology, the topic model has been used to explain a number of phenomena in semantic representation (Griffiths et al., 2007).

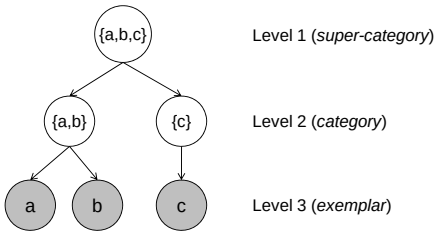


Figure 1: Illustration of our basic approach using a toy dataset with a simple tree structure and three exemplars a, b, and c, each assigned to its own leaf node.

The success of the topic model in explaining concepts from semantic representation suggests that it may also be able to explain phenomena involving feature representation.

Feature representation holds that concepts are represented as collections of (usually binary) features (Markman, 1999). For instance, the concept SNAKES would be a collection of features such as *is an animal*, *is brown*, *slithers*, *has a forked tongue* and *is dangerous*. Here we use a novel variant of the topic model, which incorporates information about category hierarchies, to learn the feature representations of a number of hierarchically related concepts.

The basic idea behind our model is that, in a hierarchy, each feature associated with a concept is either inherited from a concept of which it is an exemplar, or is idiosyncratic to the concept itself. For example, in the hierarchy ANIMALS → MAMMALS → DOGS, the features of an exemplar of a *dog* could be attributed to the fact that (1) dogs are ANIMALS—as with the feature "breathes", (2) dogs are MAMMALS—as with the feature "has hair", and (3) dogs are DOGS—such as the feature "barks". Our model uses the features assigned to exemplars of the concept DOGS and other animals to resolve both the features of the concepts at various levels of abstraction (e.g., DOGS, MAMMALS, and ANIMALS) and the degree to which each exemplar's features are inherited from each concept. In the remainder of this section, we first present a conceptual description of our model, and then formalize this in a hierarchical bayesian framework.

Conceptual Description of Approach Consider the illustration shown in Figure 1, showing a simple hierarchy of concepts. Suppose that we know the structure of this hierarchy, but are only given the sets of features for the three shaded nodes corresponding to the exemplars a, b, and c. Despite the fact that we have no information about the unshaded category nodes, intuitively we ought to be able to make reasonable guesses about their features. Specifically, we might assume that a, b and c, derive some of their features from each of their ancestors. This would mean then that the category $\{a, b\}$, the parent of exemplars a and b, possesses all of the features that are common to a and b. Additionally, the category $\{a, b, c\}$ would possess all of the features that are shared by its two children, the categories $\{a, b\}$ and $\{c\}$, the features that are shared by all of a, b, and c. This would allow us to infer feature representations for the unshaded ancestor nodes.

We need now to represent exemplars a, b, and c in a way that allows us to make use of the information contained in the hierarchy shown in Figure 1. We do this by assuming that the features we observe in an exemplar are a mixture of features from its parents and its own idiosyncratic features. For example, the features of exemplar a are a mixture of features from categories $\{a, b, c\}$ and $\{a, b\}$ and a. Modeling an exemplar in this way allows us to learn the features of the unobserved categories $\{a, b, c\}$, $\{a, b\}$, and $\{c\}$, as well as the degree to which each category contributes to the representation of each exemplar. Inference for this problem involves jointly learning featural representations for all nodes in the graph and the the mixture weights of all exemplars.

Formal Model Description In this section, we present the details of the approach we outlined in the previous section. We begin by formalizing the model in terms of the graph presented in Figure 1, and then extend this description to a model for an arbitrary graph structure. In the graph in Figure 1, we have $C = 6$ nodes: a, b, c, $\{a, b\}$, $\{c\}$ and $\{a, b, c\}$. To each of these nodes in the graph (c_i) we associate a multinomial distribution over the V unique features present in the dataset. Each exemplar in our model (d_j) is represented by a probability distribution θ_j over a subset of the C nodes in the graph. Note that each exemplar has an associated concept in the graph and that each concept has an associated node in the graph.

In order to exploit the hierarchical nature of the graph, we assume that each exemplar is a distribution over the nodes to which it was originally assigned (which is observed data), as well to *all of the ancestor nodes* of those nodes. So, for example, an exemplar d that was assigned to node a in the graph is represented by a weighted distribution (θ_d) over the node a and its two ancestor nodes, $\{a, b\}$ and $\{a, b, c\}$. Each of these three nodes is represented by a multinomial distribution $\phi_{\cdot}$ over features $x_{i=1,\ldots,V}$. Given these exemplar's distribution over nodes, as well as the nodes' distributions over features, we can express the features of exemplar d as a weighted sum of the these three nodes: $p(x_i|d) \propto \sum_{\text{nodes}} p(x_i|\phi_{\text{node}}) \times p(\text{node}|\theta_d)$

We now generalize this to an arbitrary hierarchical graph structure, where C is the number of unique nodes in the graph and the j^{th} node is represented by c_j. Each node c is a V-dimensional multinomial distribution ϕ_c over the set of V available features. For exemplar d, we observe both the vector of feature counts $\mathbf{x^{(d)}}$ as well as the initial assignments of the exemplar to one or more nodes $\mathbf{c^{(d)}}$. We extend the set of initial node assignments for exemplar d to be the set of *Assigned + Ancestor* nodes, $c^{(d)}$, where we distinguish the complete set of nodes associated with j from the observed node assignments by putting the observed set in bold. Each exemplar is associated with a multinomial distribution θ_d over $c^{(d)}$. The random vector θ_d is sampled from a Dirichlet distribution with hyper-parameter $\boldsymbol{\alpha}^{(d)}$, where $\boldsymbol{\alpha}^{(d)}$ is a vector with dimension equal to the number of nodes in the set $c^{(d)}$.

Given a hierarchical graph structure, and the set of observed features and node-assignments for each exemplar, the

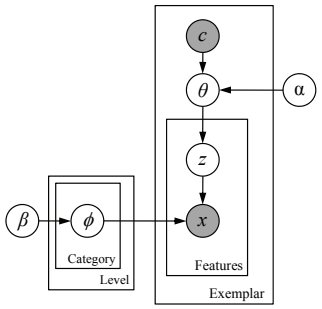


Figure 2: Model illustrated using graphical model notation

generative process for this model is:

1. For each node $c \in \{1, \ldots, C\}$, sample a multinomial distribution over feature-types $\phi_c \sim \text{Dirichlet}(\cdot|\beta)$
2. For each exemplar $d \in \{1, \ldots, D\}$
   (a) Sample a multinomial distribution over the set of nodes $c^{(d)}$, $\theta_d \sim \text{Dirichlet}(\cdot|\alpha^{(d)})$
   (b) For each feature $i \in \{1, \ldots, N_d^X\}$
      i. Sample a node $z_i \sim \text{Discrete}\,(\theta_d)$
      ii. Sample a feature $x_i^{(d)} \sim \text{Discrete}\,(\phi_{z_i})$ from the node $c = z_i$

This model is presented using graphical model notation in Figure 2.

Experiments

We applied our model to two datasets: (1) A set of animal-feature matrices from the Leuven Concept Database, and (2) a set of documents extracted from a subgraph of the Wikipedia category structure. Despite the very different nature of these datasets, the model is perfectly applicable in both cases. However, for clarity, we describe below two of the major differences between these datasets.

Datasets In the Leuven Animal Concept Database, features are counts of *animal features* from an Animal $\times$ Feature matrix [1]. In the case of the Wikipedia dataset, these features are counts of *words* from a Document $\times$ Word matrix. In the Leuven Concept Database, the exemplars correspond to the 129 unique animals in the dataset, whereas in the Wikipedia dataset, the exemplars correspond to the $10,432$ documents in our dataset. Furthermore, exemplars in the Wikipedia dataset could be initially assigned to one or more of the nodes in the graph, whereas exemplars in the Leueven concept database are put in 1-1 correspondence with nodes representing specific animals (where there is only one exemplar per node). Note that this 1-1 correspondence between exemplars and nodes–although a notable distinction–does not comprise a fundamental difference between this dataset and the Wikipedia dataset. In fact, one could easily imagine a situation in which we have multiple exemplars assigned to some of the animal species in this graph (e.g., for the node *dogs*, we could have some people provide judgments about the features with respect to the breed *Rottweiler* and others provide judgments about the breed *Chihuahuas*). Even without this, it is certainly the case that each of the four subjects who provided feature judgements had slightly different representations of each animal species, and we could have used an alternative representation of the data in which each individual subject's judgments were treated as exemplars (but for simplicity chose instead to use the sums across participants).

An additional substantive difference between the datasets is in their corresponding graph structures. The Leuven Concept Database of animals can be represented by a simple tree structure with a single root node representing the broad category ANIMALS. This root node has a directed edge pointing to each of the five animal-categories (e.g., MAMMALS), and each of these five animal-categories has directed edges pointing to multiple species within those categories (e.g., DOGS)[2]. As with the Leuven Dataset, the Wikipedia dataset we used has a single root note: POLITICS BY ISSUE. However, the category structure is significantly more convoluted, and contains 361 categories with a a much wider range of subject matter and conceptual specificity across these categories (ranging, e.g., from the broad categories MILITARY, and HUMAN RIGHTS to the highly specific categories ANTI-WAR SONGS and TRANSGENDER LAW). Our approach is nonetheless directly applicable to both datasets.

Applying Model to Animal by Feature Matrices

We applied our model to the Type II matrices of the Leuven Concept Database. An illustration of these results is provided in Figure 3. The model learns a probability distribution over features for all exemplars (i.e., leaf nodes) in the database, as well as for the five Animal-Category distributions (e.g., Mammals) and the root node, "All Animals". The top eight most likely features learned by the model are shown for all of the category-level and the root-level nodes. Due to space constraints, we do the same for only six of the 129 total animal-level distributions that were learned.

Note that there was *no* observed data for the category-level or root nodes. These distributions were all learned by the model by assigning the common features among child nodes to the parent nodes. Note that these Category-level representations are quite easily interpretable, and in fact (for the most part) provide excellent definitions of these classes of animals. For example, in four out of five of the category-level distributions, the feature that defines the category itself (e.g., "is a bird"), is among the most likely features at the category level. And, even ignoring these *definition* features, the distributions are typically the standard lists of what we are taught about the categories in general (e.g., for Birds, the fact that they have wings, two feet, a bill, lay eggs, and have feathers).

[1] Note that although the elements of the Animal $\times$ Feature matrix are often treated as bernoulli probabilities, the dataset itself actually consists of *counts*, corresponding to the the number of times each feature was assigned to each animal across four participants.

[2] Although the Leuven Concept Database does not *explicitly* provide this graph structure; instead it provides five disjoint two-level trees with animal-categories as the roots and species as the leaves. However, it is implied that the animal categories can all be treated as sub-trees within an overall graph for ANIMALS

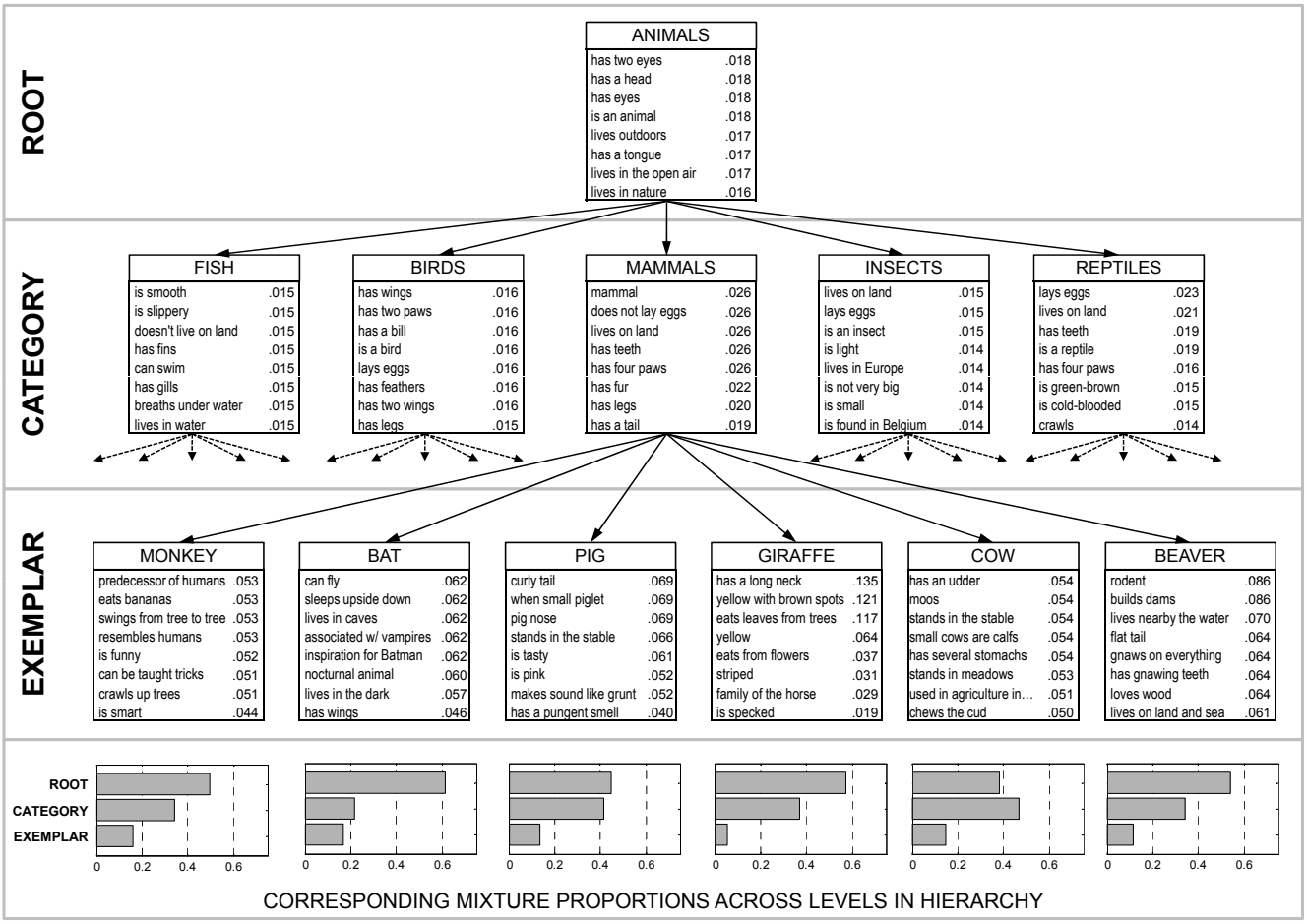


Figure 3: Illustration of our model applied to *Animal* × *Feature* matrices from the Leuven Concept Database. The eight highest-probability features are shown for the root node ANIMALS, all second-Level *Animal-Category* nodes, and six exemplars from the category MAMMALS. Below each exemplar we present its probability distribution over levels in the graph.

The only case in which the *definition*-feature doesn't appear in the top eight features is for the category FISH, in which the feature "is a fish" was the twelfth most likely feature learned by the model. Interestingly, this may be due to the fact that there were numerous misclassifications of water-mammals (specifically, Dolphins, Whales, and Orcas), to the Fish category. Thus, because several of the exemplars used to infer the features for fish were not fish, but did have many fish-like features (such as "is smooth" and "doesn't live on land"), these were the features that were pushed to the top.

We show a subset of six of the exemplars for the "Mammals" category. You can see that the category-level features, which are shared amongst all Mammals, do not appear with high-probability for the exemplar level distributions. This is because the common features are explained away (and captured at the category-level distribution for Mammals). Instead, the features that are highly likely are the features which best distinguish the exemplars from other mammals. For example, the distribution for *bat* puts high probability on many features relating to the fact that it is an unusual case of a flying mammal. What these exemplar-level distributions intuitively capture are features that might be most informative hints in a guessing game, *conditioned* on the fact that the guesser already knows the fact that the animal is a mammal.

Relationship Between Model Representation and Animal Typicality The general purpose of the previous experiment was to examine some of appealing features of modeling concepts using a distributed representation across a graph hierarchy. Namely, (1) that this approach can be used to generalize from specific exemplars to higher-level categorical representations, and (2) that it increases the specificity of the features represented at lower levels of the hierarchy by explaining away common features to higher categories. This approach was not conceived directly as a means to predict additional types of data, such as similarity ratings or typicality ratings. However, if our approach is to provide a useful framework for understanding how people represent categories, it is important to connect it with such types of data (for this paper, we restrict our analysis to typicality ratings).

One thing which falls directly out of the model is the extent to which each animal provides a good representation of each category. Specifically, the relative probability of the category-level node, given each animal exemplar, provides a

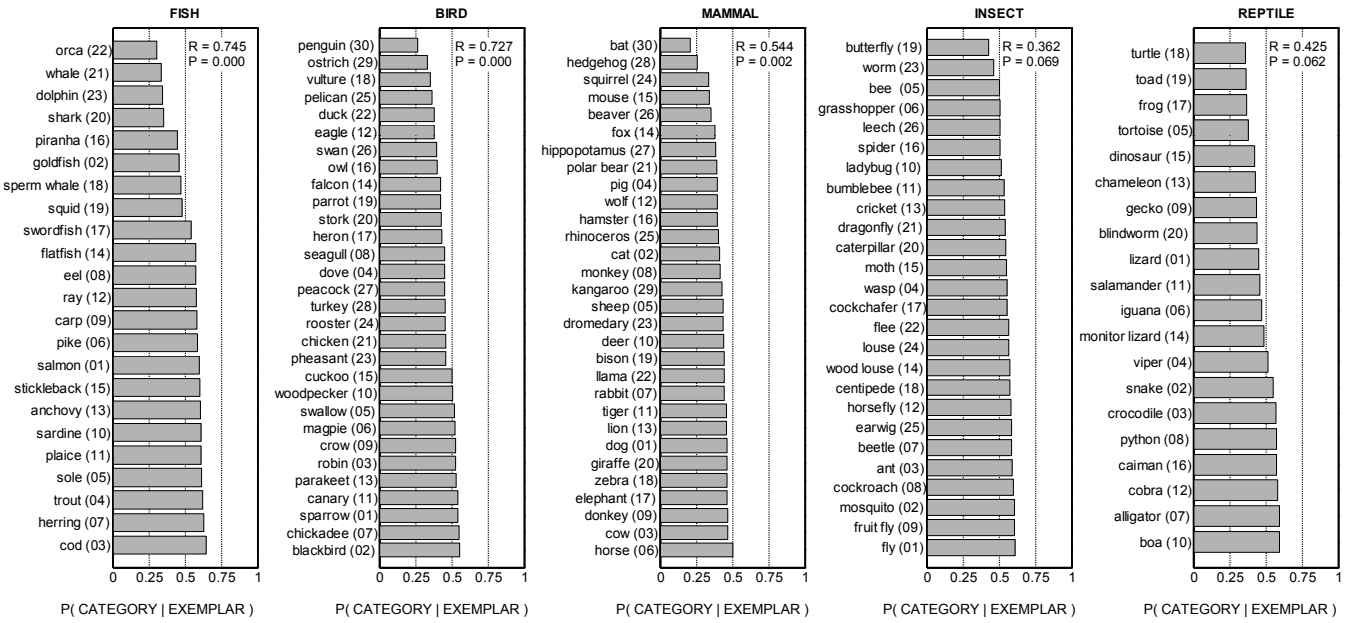


Figure 4: All animal exemplars and category-goodness rankings (in parentheses) for each animal-category, sorted according to the $p(Category|Exemplar)$ assigned by our Model.

natural measure of how typical the animal is of the category. To compare this with human judgments, we used the "goodness rankings" of each animal, which was collected as part of the Leuven Concept Database. For our analysis, we averaged across the rankings of the 20 participant rankings within the database to create a single ranking, and then rescaled all values from zero to one so that all categories of animals would have the same range of scores. We then compared these values with the mixture weights that the model assigned to each exemplar at the ANIMAL CATEGORY level of the hierarchy (i.e., the $p(\text{Category}|\text{Exemplar})$).

The relationship between the $p(\text{Category}|\text{Exemplar})$ and the typicality scores is shown in Figure 4. For each category, we provide a list of all animals and their corresponding (unscaled) goodness rankings, sorted by increasing $p(\text{Category}|\text{Exemplar})$ learned by the model. By visual inspection, one can see that atypical animals (those with lower rankings) are assigned less weight by the model than typical animals. For example, *Penguins*, were ranked as the least typical animal in the BIRD category, are assigned by far the least weight by the model at the category level. The most highly weighted birds, *blackbirds*, *chicadees*, and *sparrows*, were rated second, seventh, and first most typical out of the thirty birds in the dataset.

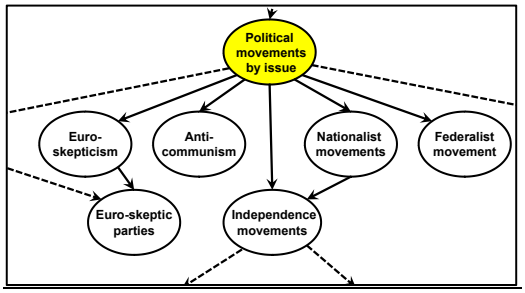
To provide a qualitative measure of how well the model predictions corresponded to human typicality rankings, we computed the R^2 statistic to measure the correlation between the $p(\text{Category}|\text{Exemplar})$ and the goodness scores within each category. The correlations were highly significant for three of the categories ($p < .001$ for BIRDS and FISH, and $p = .002$ for a MAMMALS), and nearly significant at the $\alpha = .05$ level for the INSECT and REPTILE categories ($p = .069$ and $p = .062$, respectively).

One interesting note is that four water-mammals were actually misclassified in the Leuven dataset as FISH. Notably, the model picks up on these misclassifications quite well; the three least-weighted animals by the model were all in fact examples of these misclassified mammals (*dolphins*, *whales*, and *orcas*). Furthermore, the model captures the misclassifications quite well in terms of its featural representations; the three highest-probability features learned at the *exemplar* level for all three of these animals was "mammal". The reason for this is that when the "mammal" feature is assigned to an animal that is not in the MAMMAL category, this feature cannot be "explained away" by any of its ancestor nodes (because the feature "mammal" will have a very low probability in the category-level representations for all non-mammal categories, as well as for the root category ANIMALS). One implication of these results is that our model may be useful for capturing misclassifications in an ontology.

Applying The Model to Wikipedia Documents

To demonstrate that the model we describe in this paper is applicable to real-world datasets, where the categories are less carefully constructed and features are much noisier, we applied our model to a set of documents from the subset of the Wikipedia category structure (described previously). In the Wikipedia dataset, each exemplar is a *document*, and the features of each document are the *word-counts* for that document. Our Wikipedia dataset had 361 concept nodes, where the root-node was POLITICS BY ISSUE, and $10,432$ documents which could be assigned to one or more categories.

We present two main results below, showing (1) that the model is able to generalize to nodes for which there is no directly-assigned data and learn a reasonable feature representation for these nodes, and (2) that the model improves the



Political movements by issue	.010	Political organizations by issue	.010	Political parties by issue	.010
party	.017	member	.029	party	.104
war	.012	group	.027	election	.066
communist	.012	organization	.015	political	.031
movement	.011	united	.014	parties	.028
national	.011	government	.012	seat	.021
government	.010	support	.009	democratic	.015
member	.010	organisation	.009	parliament	.013
political	.010	police	.009	parliamentary	.013
anti	.009	french	.007	coalition	.012
soviet	.008	attack	.007	union	.011
union	.007	founded	.007	vote	.011
german	.007	political	.006	votes	.011
germany	.006	campaign	.006	won	.010
organization	.006	leader	.006	general	.009
fascist	.006	military	.005	member	.008
socialist	.006	terrorist	.005	led	.008
nazi	.006	states	.005	polish	.008

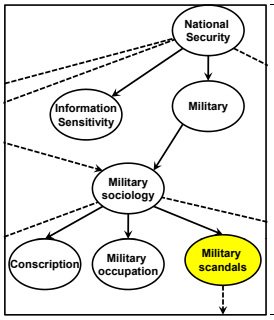
Figure 5: **Top**: A small subgraph of our Wikipedia dataset. **Bottom**: The most most likely features learned for three categories which had not been directly assigned any documents.

specificity of these feature representations when compared with a "flat" version of the model which does not account for the graph hierarchy [3].

Generalization to nodes with no data Figure 5 illustrates the ability of our model to generalize to nodes with missing data. In the top panel, we show a small portion of the subgraph, highlighting a node to which no documents were assigned. However, since there were many descendants of this node which contained data, common words from these descendants were explained away to this node. The bottom panel of this figure shows the most likely words for this node and two additional nodes which had no documents directly assigned to them. Looking at these distributions, one can see that the model comes up with reasonable distributions over features for each of these nodes.

Leveraging Graph Structure to Improve Category Specificity Figure 6 illustrates the effect of allowing features to be assigned to ancestors of nodes to which they are assigned. In the left panel of this figure, we show a relatively dense region near the lower levels of the Wikipedia graph, containing the category MILITARY SCANDALS (highlighted). In the right panel of this figure, we compare the distribution learned for the "Flat" version of our model—which only assigns probability to *observed* category-assignments—compared to the distribution learned by the graph-based model. Note that the high-probability words learned by the graph-based model are much more specific to the *scandals* aspect of this category, while the "flat" model has many more words associated with the military *in general*. In the graph-based model, these

[3] In the "flat" version of the model, features can only be assigned to the set of *observed* labels for each document, rather than to the set of both *assigned labels* plus *ancestor labels*.



FLAT LDA MODEL "Military Scandals"	.010	GRAPH-BASED LDA MODEL "Military Scandals"	.002
military	.007	army	.015
war	.007	military	.013
air	.005	french	.011
army	.005	torture	.010
states	.005	general	.010
united	.005	gladio	.009
report	.004	officer	.008
attack	.004	prisoner	.008
soldier	.004	dreyfus	.008
government	.003	investigation	.008
japanese	.003	soldier	.008
general	.003	massacre	.007
soviet	.003	minister	.007
force	.003	report	.007
new	.003	commission	.007
member	.003	abu	.007
officer	.003	ghraib	.007

Figure 6: Comparison of our model representation of the category MILITARY SCANDALS with a similar model that does not account for graph structure.

more general words tend to be assigned further up the hierarchy (specifically, these words will be drawn to the more general category MILITARY, which is an ancestor of MILITARY SCANDALS).

Conclusions

This paper presented a novel model for representing exemplar features using distributed representations across a hierarchical graph structure. Using data consisting of Animal by Feature matrices, we demonstrated that this model infer reasonable featural representations for higher-level categories, by generalizing from the features present amongst the exemplars of a category. We furthermore showed that the inferred representation of *species-level* exemplars at the *animal-category* level of abstraction closely corresponds to people's judgements about how representative a species is of a category. Finally, using our Wikipedia dataset we demonstrated that this model can similarly perform *generalization* in a much noisier, real-world context, as well as improve the specificity of its featural representation of categories over similar models which do not account for category hierarchies. In future work, we will explore whether the model can contribute to the understanding of additional psychological data such as similarity ratings.

Acknowledgements We thank our three anonymous reviewers for their helpful comments.

Références

Austerweil, J., & Griffiths, T. L. (2009). Analyzing human feature learning as nonparametric Bayesian inference. In *Advances in neural information processing systems 21*.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003, January). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*, 240-248.

de Deyne, S., Verheyen, S., Ameel, E., Vanpaemel, W., Dry, M., & Voorspoels, W. (2008). Exemplar by feature applicability matrices and other Dutch normative data for semantic concepts. *Behavior Research Methods*, *40*(4), 1030-1048.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, *114*(2), 211-244.

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences*, *105*(31), 10687-10692.

Kemp, C., & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review*.

Markman, A. B. (1999). *Knowledge representation*. Mahwah, NJ : Lawrence Erlbaum Associtates.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: studies in the internal structure of categories. *Cognitive Psychology*, *7*.

Shepard, R. N. (1980, 24 octobre). Multidimensional scaling, tree-fitting, and clustering. *Science*, *210*(4468), 390–398.

Zeigenfuse, M. D. (2010). *Feature importance in mental representation*. Thèse de doctorat non publiée, University of California, Irvine.