A Bayesian Model for Joint Learning of Categories and their Features

Lea Frermann and Mirella Lapata
l.frermann@ed.ac.uk, mlap@inf.ed.ac.uk

School of Informatics
Institute for Language, Cognition, and Computation
The University of Edinburgh

June 3rd, 2015

"Cognition is Categorization" (Harnad, 2005)

The ability to generalize from experience

- underlying a variety of common mental tasks, such as learning, perception or language use

"Cognition is Categorization" (Harnad, 2005)

The ability to generalize from experience

- underlying a variety of common mental tasks, such as learning, perception or language use

Learning categories

Is a glove a piece of clothing?

Learning features

Do all pieces of clothing have stripes?

Concepts, Categories and Features

learning categories $\leftrightarrow$ **learning features**

(Schyns and Rodet, 1997; Goldstone et al., 2001)

Category

Features

- wool
- leather
- is dotted
- has color
- keeps warm
- keeps dry
- is fashionable

Concepts, Categories and Features

learning categories $\leftrightarrow$ **learning features**

(Schyns and Rodet, 1997; Goldstone et al., 2001)

Category

selected by freepik.com

Features

- wool
- leather
- is dotted
- has color
- keeps warm
- keeps dry
- is fashionable

Concepts

Concepts, Categories and Features

learning categories $\leftrightarrow$ **learning features**

(Schyns and Rodet, 1997; Goldstone et al., 2001)

Category

Features

- wool
- leather
- is dotted
- has color
- keeps warm
- keeps dry
- is fashionable

Concepts

Features and Feature Types

Features are structured

(McRae et al., 2005; Spalding and Ross, 2000)

Category

selected by freepik.com

Features

- wool
- leather
- is dotted
- has color
- keeps warm
- keeps dry
- is fashionable

Features and Feature Types

Features are structured

(McRae et al., 2005; Spalding and Ross, 2000)

Category

selected by freepik.com

Features

- wool
- leather
- is dotted
- has color
- keeps warm
- keeps dry
- is fashionable

Feature Types

← **material**

← **appearance**

← **function**

Features and Feature Types

Features are structured

(McRae et al., 2005; Spalding and Ross, 2000)

Category

Features

- wool
- leather
- is dotted
- has color
- keeps warm
- keeps dry
- is fashionable

Feature Types

← **material**

← **appearance**

← **function**

The feature type distribution varies across categories.

(Ahn, 1998)

Contributions

(I) the first joint model of category and feature acquisition

- principled formulation in the Bayesian framework
- knowledge-lean
- large-scale modeling and evaluation
  $\rightarrow$ learning from textual input

(II) a way of quantitatively evaluating the learnt features

- avoiding direct comparison with human-produced feature sets

Related Work

Related models

- text-based category acquisition
  (Fountain and Lapata, 2011; Frermann and Lapata, 2014)
- highly engineered feature extraction from text
  (Baroni et al., 2010; Kelly et al., 2014)

- small-scale experiments (Anderson, 1991; Sanborn et al., 2006)
- artificial stimuli

Feature evaluation

- comparison of text-based features to human-produced feature sets (problematic)
  (Baroni et al., 2010; Kelly et al., 2014)

Learning Objectives

1. accurate semantic categories

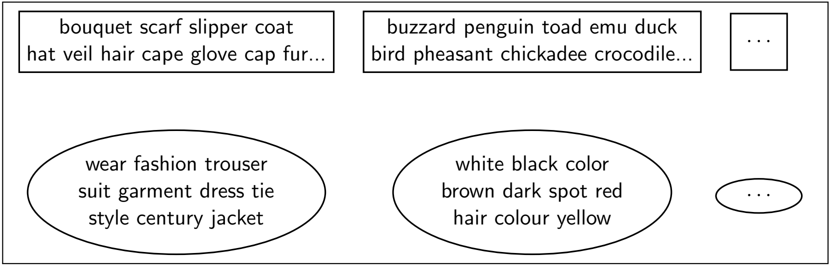
bouquet scarf slipper coat
hat veil hair cape glove cap fur...

buzzard penguin toad emu duck
bird pheasant chickadee crocodile...

...

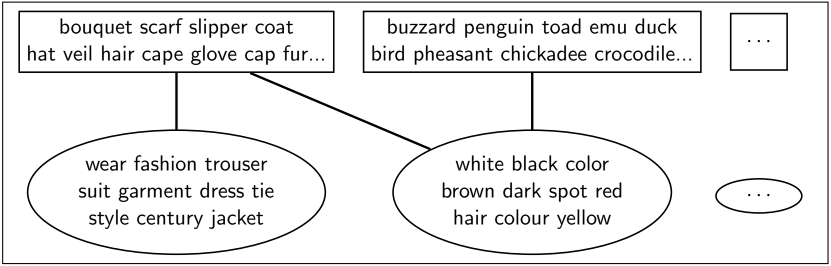
Learning Objectives

1. accurate semantic categories
2. coherent feature types



Learning Objectives

1. accurate semantic categories
2. coherent feature types
3. relevant category-feature type associations



Input

Textual input (documents) from text corpora

- one **target concept**
- sentence context as features

… a	**skirt**	is a tube- or cone-shaped garment …
… a	**skirt**	covers all or part of the legs …
…	**skirts**	are more commonly worn by women …

Proxy of the cognitive learning environment

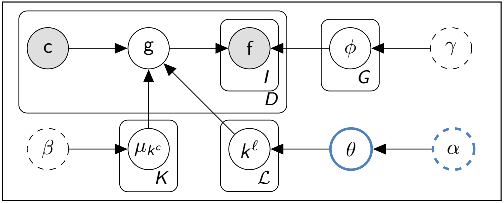
examples adapted from `https://en.wikipedia.org/wiki/Skirt`.

Model Overview

Assumptions / Modeling Decisions

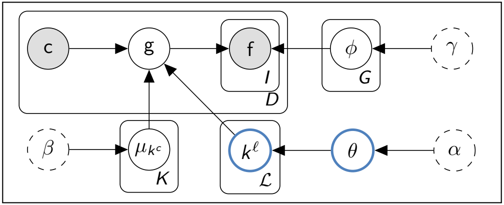
- one category k per concept type
  (hard clustering of concepts into categories)
- one feature type g per input
  (soft clustering of features into feature types)
- feature types capture one particular aspect of meaning
- categories differ in their feature type associations

Generative Story



Draw distribution over categories $\theta \sim Dir(\alpha)$

Generative Story

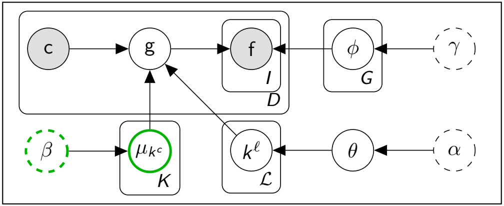


Draw distribution over categories $\theta \sim Dir(\alpha)$

for concept type ℓ **do**

  Draw category $k^\ell \sim Mult(\theta)$

Generative Story

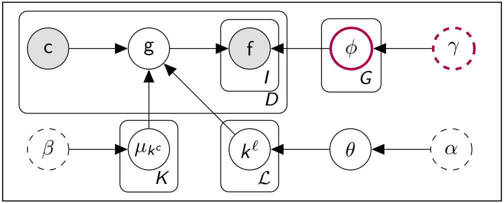


Draw distribution over categories $\theta \sim Dir(\alpha)$

for concept type ℓ **do**
  Draw category $k^\ell \sim Mult(\theta)$

for category k **do**
  Draw feature type distribution $\mu_k \sim Dir(\beta)$

Generative Story



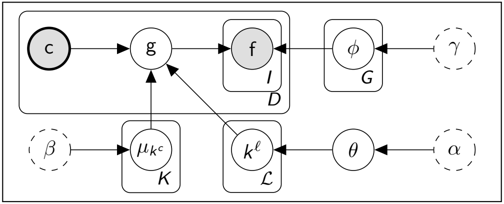
Draw distribution over categories $\theta \sim Dir(\alpha)$

for concept type ℓ **do**
  Draw category $k^\ell \sim Mult(\theta)$

for category k **do**
  Draw feature type distribution $\mu_k \sim Dir(\beta)$

for feature type g **do**
  Draw feature distribution $\phi_g \sim Dir(\gamma)$

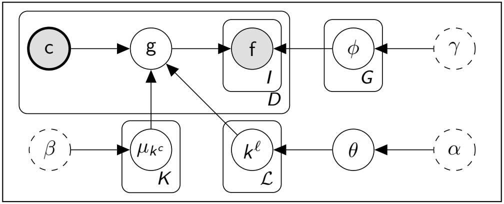
Generative Story



distribution over categories $\theta \sim Dir(\alpha)$
for concept type ℓ **do**
    category $k^\ell \sim Mult(\theta)$
for category k **do**
    feat type distribution $\mu_k \sim Dir(\beta)$
for feature type g **do**
    feature distribution $\phi_g \sim Dir(\gamma)$

for input d **do**

Generative Story



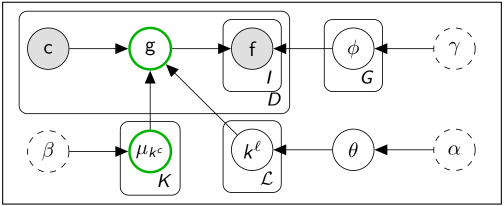
distribution over categories $\theta \sim Dir(\alpha)$
for concept type ℓ **do**
  category $k^\ell \sim Mult(\theta)$
for category k **do**
  feat type distribution $\mu_k \sim Dir(\beta)$
for feature type g **do**
  feature distribution $\phi_g \sim Dir(\gamma)$

for input d **do**

  Observe concept c^d and retrieve category k^{c^d}

clothing → **skirt**

Generative Story



distribution over categories $\theta \sim Dir(\alpha)$
for concept type ℓ **do**
  category $k^\ell \sim Mult(\theta)$
for category k **do**
  feat type distribution $\mu_k \sim Dir(\beta)$
for feature type g **do**
  feature distribution $\phi_g \sim Dir(\gamma)$

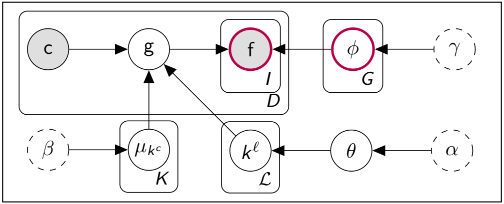
for input d **do**

  Observe concept c^d and retrieve category k^{c^d}

  Generate feature type $g^d \sim Mult(\mu_{k^{c^d}})$

clothing $\rightarrow$ appearance $\rightarrow$ **skirt**

Generative Story



distribution over categories $\theta \sim Dir(\alpha)$
for concept type ℓ **do**
  category $k^\ell \sim Mult(\theta)$
for category k **do**
  feat type distribution $\mu_k \sim Dir(\beta)$
for feature type g **do**
  feature distribution $\phi_g \sim Dir(\gamma)$

for input d **do**

  Observe concept c^d and retrieve category k^{c^d}

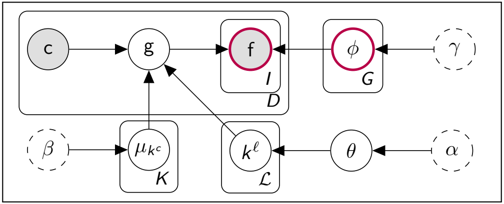
  Generate feature type $g^d \sim Mult(\mu_{k^{c^d}})$

  for context position $i = \{1..I\}$ **do**

    draw context word $f_{d,i} \sim Mult(\phi_{g^d})$

clothing $\rightarrow$ appearance $\rightarrow$ blue **skirt**

Generative Story



distribution over categories $\theta \sim Dir(\alpha)$
for concept type ℓ **do**
category $k^\ell \sim Mult(\theta)$
for category k **do**
feat type distribution $\mu_k \sim Dir(\beta)$
for feature type g **do**
feature distribution $\phi_g \sim Dir(\gamma)$

for input d **do**

Observe concept c^d and retrieve category k^{c^d}

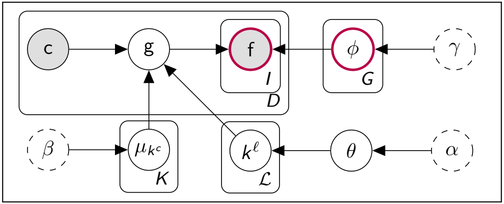
Generate feature type $g^d \sim Mult(\mu_{k^{c^d}})$

for context position $i = \{1..I\}$ **do**

draw context word $f_{d,i} \sim Mult(\phi_{g^d})$

clothing $\rightarrow$ appearance $\rightarrow$ blue **skirt** sprinkled

Generative Story



distribution over categories $\theta \sim Dir(\alpha)$

for concept type ℓ **do**

  category $k^\ell \sim Mult(\theta)$

for category k **do**

  feat type distribution $\mu_k \sim Dir(\beta)$

for feature type g **do**

  feature distribution $\phi_g \sim Dir(\gamma)$

for input d **do**

  Observe concept c^d and retrieve category k^{c^d}

  Generate feature type $g^d \sim Mult(\mu_{k^{c^d}})$

  for context position $i = \{1..I\}$ **do**

    draw context word $f_{d,i} \sim Mult(\phi_{g^d})$

clothing $\rightarrow$ appearance $\rightarrow$ blue **skirt** sprinkled red

Learning with Blocked Gibbs Sampling

1. For each **input**: re-sample feature type
   - fix the current concept categorization
   - establish meaningful feature types

$$\boxed{\begin{aligned}\propto\ & p(\mathit{featuretype} = i|\mathit{category}(\mathit{concept}), ...) \times \\ & p(\mathit{features}|\mathit{featuretype} = i, ...)\end{aligned}}$$

Learning with Blocked Gibbs Sampling

1. For each **input**: re-sample feature type
   - fix the current concept categorization
   - establish meaningful feature types

$$\propto \; p(featuretype = i | category(concept), \ldots) \times$$
$$p(features | featuretype = i, \ldots)$$

2. For each **concept**: re-sample category
   - fix the current feature types
   - group concepts with similar feature type associations into the same category

$$\propto \; p(category(concept) = j | \ldots) \times$$
$$p(featuretypes(concept) | category(concept) = j, \ldots)$$

Evaluation: Overview

Models

- Strudel – pattern-based feature extraction from text
  (Baroni et al., 2010)
- BayesCat – a Bayesian model of category acquisition
  (Frermann and Lapata, 2014)
- BCF – this work ; "Bayesian Categories and Features"

Evaluation: Overview

Models

- **Strudel – pattern-based feature extraction from text**

  (Baroni et al., 2010)

  - feature extraction through pattern matching in the context of a concept's occurrence
  - Strudel learns features for concepts
    categories and feature types are constructed post-hoc

- BayesCat – a Bayesian model of category acquisition

  (Frermann and Lapata, 2014)

- BCF – this work ; "Bayesian Categories and Features"

Evaluation: Overview

Models

- Strudel – pattern-based feature extraction from text
  (Baroni et al., 2010)
- **BayesCat – a Bayesian model of category acquisition**
  (Frermann and Lapata, 2014)
  - features are learnt as a by-product
  - no feature types are learnt
- BCF – this work ; "Bayesian Categories and Features"

Evaluation: Overview

Models

- Strudel – pattern-based feature extraction from text
  (Baroni et al., 2010)
- BayesCat – a Bayesian model of category acquisition
  (Frermann and Lapata, 2014)
- **BCF – this work ; "Bayesian Categories and Features"**

Evaluation: Overview

Models

- Strudel – pattern-based feature extraction from text
  (Baroni et al., 2010)
- BayesCat – a Bayesian model of category acquisition
  (Frermann and Lapata, 2014)
- BCF – this work ; “Bayesian Categories and Features”

Data

- standard set of categorized basic-level target concepts
  (McRae et al., 2005; Vinson and Vigliocco, 2008; Fountain and Lapata, 2010)
- input documents extracted from Wikipedia

Evaluation: Overview

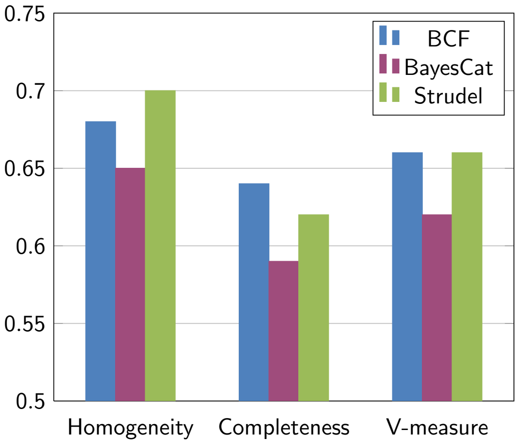
Models

- Strudel – pattern-based feature extraction from text
  (Baroni et al., 2010)
- BayesCat – a Bayesian model of category acquisition
  (Frermann and Lapata, 2014)
- BCF – this work ; "Bayesian Categories and Features"

Questions

1. How meaningful are the learnt categories?
2. How predictive are the features?
3. How internally coherent are the learnt feature types?
4. How meaningful are the category-feature type associations?

How meaningful are the learnt categories?



- Strudel is highly optimized, categories formed post-hoc
- BCF performance competitive with Strudel's

How predictive are the features?

Given the context (features) of an unseen document: predict its target concept

Unseen input document

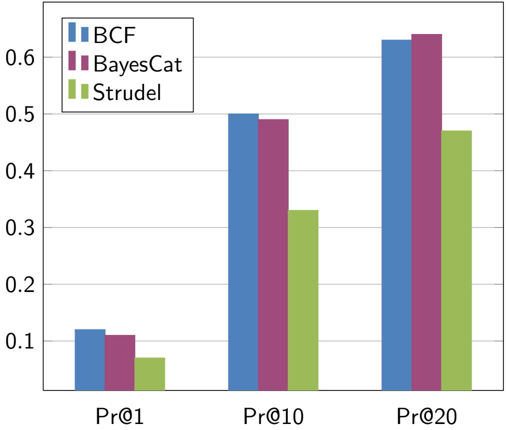
salmon	`journey move hundred mile strong current reproduce`

Model Predictions

BCF	**salmon**	tuna	goldfish	lobster	fish
BayesCat	fish	radio	goldfish	**salmon**	clock
Strudel	train	house	apartment	ship	car

How predictive are the features?

Given the context (features) of an unseen document: predict its target concept



- patterns may be too restrictive, feature set too constrained (high precision, low recall)

How coherent are the learnt feature types?

Text-elicited feature types are not compatible with human elicited ones (Baroni et al., 2010; Kelly et al., 2014)

We evaluate feature types directly through crowdsourcing (AMT)

Topic Intrusion Paradigm (Chang et al., 2009)

- given a list of topic words, spot the inserted 'intruder word'
- easy if topics are coherent

How coherent are the learnt feature types?

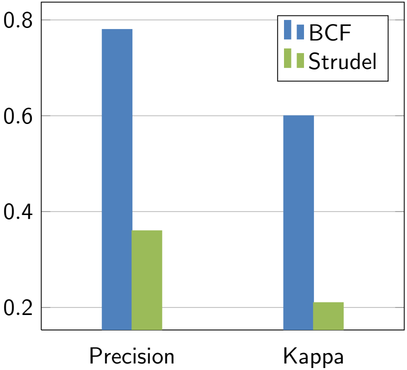
"Select the intruder word."

○	○	●	○	○	○
egg	female	box	young	bird	food

How coherent are the learnt feature types?

“Select the intruder word.”

○	○	●	○	○	○
egg	female	box	young	bird	food



How coherent are the learnt feature types?

Example Output for category **clothing**

Strudel

- white change pant trouser replace
- paint fasten thick layer apply
- hole shirt lie neck finish
- pattern hood cover crimson woolen
- man occasion see steal striped

BCF

- wear cover veil woman coat
- white black color brown dark
- eye tooth ear skin lip
- wear suit trouser woman garment
- animal feather skin wool material

Conclusion

Summary

- the first model for joint learning of categories and features
- principled formulation in one unified framework
- competitive results compared with an engineered, pipelined feature extraction system
- feature evaluation method that sidesteps comparison to human produced features

In the future

- multimodal learning
- incremental learning (e.g. particle filtering)
- dynamic models: tracking meaning change over time

Conclusion

Summary

- the first model for joint learning of categories and features
- principled formulation in one unified framework
- competitive results compared with an engineered, pipelined feature extraction system
- feature evaluation method that sidesteps comparison to human produced features

In the future

- multimodal learning
- incremental learning (e.g. particle filtering)
- dynamic models: tracking meaning change over time

Thank you!

References I

Woo-Kyoung Ahn. 1998. Why are different features central for natural kinds and artifacts?: the role of causal status in determining feature centrality. *Cognition*, 69:135.

John R. Anderson. 1991. The adaptive nature of human categorization. *Psychological Review*, 98:409–429.

Joseph L. Austerweil and Thomas L. Griffiths. 2009. Analyzing human feature learning as nonparametric Bayesian inference. In *Advances in Neural Information Processing Systems (NIPS)*.

Marco Baroni, Brian Murphy, Eduard Barbu, and Massimo Poesio. 2010. Strudel: A corpus-based semantic model based on properties and types. *Cognitive Science*, 34(2):222–254.

Jonathan Chang, Jordan Boyd-Graber, Chong Wang, Sean Gerrish, and David M. Blei. 2009. Reading tea leaves: How humans interpret topic models. In *Neural Information Processing Systems*, pages 288–296.

References II

Trevor Fountain and Mirella Lapata. 2010. Meaning representation in natural language categorization. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, pages 1916–1921. Portland, Oregon.

Trevor Fountain and Mirella Lapata. 2011. Incremental models of natural language category acquisition. In *Proceedings of the 33nd Annual Conference of the Cognitive Science Society*, pages 255–260. Boston, Massachusetts.

Lea Frermann and Mirella Lapata. 2014. Incremental Bayesian learning of semantic categories. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2014, April 26-30, 2014, Gothenburg, Sweden*, pages 249–258.

Robert L. Goldstone, Yvonne Lippa, and Richard M. Shiffrin. 2001. Altering object representations through category learning. *Cognition*, 78:27–43.

References III

Stevan Harnad. 2005. To cognize is to categorize: Cognition is categorization. In Claire Lefebvre and Henri Cohen, editors, *Handbook of Categorization. Summer Institute in Cognitive Sciences on Categorisation*. Elsevier.

Colin Kelly, Barry Devereux, and Anna Korhonen. 2014. Automatic extraction of property norm-like data from large text corpora. *Cognitive Science*, 38(4):638–682.

Ken McRae, George S. Cree, Mark S. Seidenberg, and Chris McNorgan. 2005. Semantic feature production norms for a large set of living and nonliving things. *Behavioral Research Methods*, 37(4):547–59.

Adam N. Sanborn, Thomas L. Griffiths, and Daniel J. Navarro. 2006. A more rational model of categorization. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, pages 726–731. Vancouver, Canada.

References IV

Philippe G Schyns and Luc Rodet. 1997. Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23:681–696.

T. L. Spalding and B. H. Ross. 2000. Concept learning and feature interpretation. *Memory & Cognition*, 28:439–451.

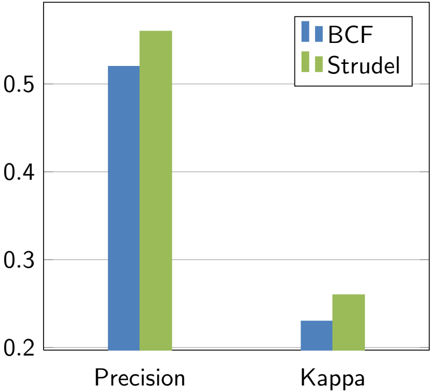
David Vinson and Gabriella Vigliocco. 2008. Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*, 40(1):183–190.

How meaningful are the feature types?

“Select intruder feature type (right) wrt category (left).”

ant hornet	○ egg female food young bird
butterfly moth	○ ant insect butterfly wasp larva
flea beetle	● wear cover veil woman coat
grasshopper	○ body air fish blood muscle
wasp caterpillar	○ sound human nerve bird brain
cockroach	○ culture symbol popular feature animal

How meaningful are the feature types?



- Strudel slightly better (differences not statistically significant)
- simpler BCF model learns qualitatively comparable features