Incremental Bayesian Learning of Semantic Categories

Lea Frermann
l.frermann@ed.ac.uk

Mirella Lapata
mlap@inf.ed.ac.uk

What is Category Learning and Why Do We Care?

Categorization is the process by which people group exemplars into categories and use those categories to reason about new exemplars they encounter.

- Category learning underlies a variety of common mental tasks including perception, learning, and the use of language.
- We focus on natural categories (ANIMALS, INSTRUMENTS, CLOTHES,...)
- We approximate the learning environment by large corpora
- We learn categories and their features incrementally in one process

Contributions

Traditionally

- Small-scale experiments with hand-coded features (Anderson (1991))
- Often artificial categories and features (Sanborn et al (2006))
- Feature Norms as proxy for human category representation (McRae et al (2005))

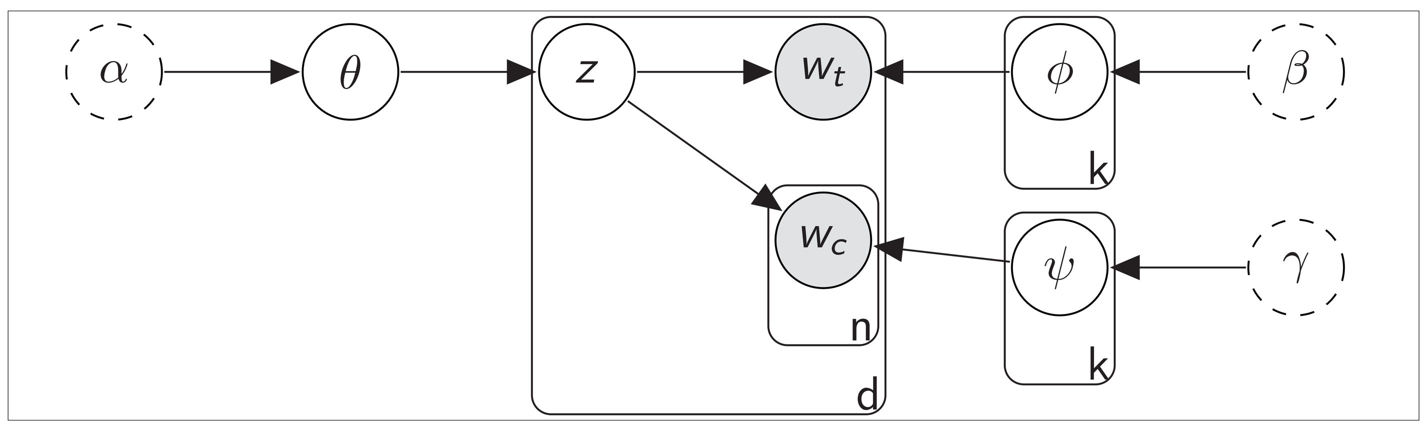
FRUIT	apple,pear	*is_eadible, is_healthy, has_seeds, is_grown*
FURNITURE	chair, desk	*found_in_home, comfortable, made_of_wood*

This work

- Language as proxy for human category representation (categories are latent)
- Potential for large-scale experiments with real-world categories (>1M observations, 40 categories, 550 concepts, 8K features)

c_{-n}	c_{-n-1}	…	c_{-1}	t	c_{+1}	c_{+2}	…	c_{+n}
		grow	*tree*	apple	*sweet*	*taste*		
		kitchen	*table*	chair	*sit*	*breakfast*		

The Bayesian Categorization Model (BayesCat)



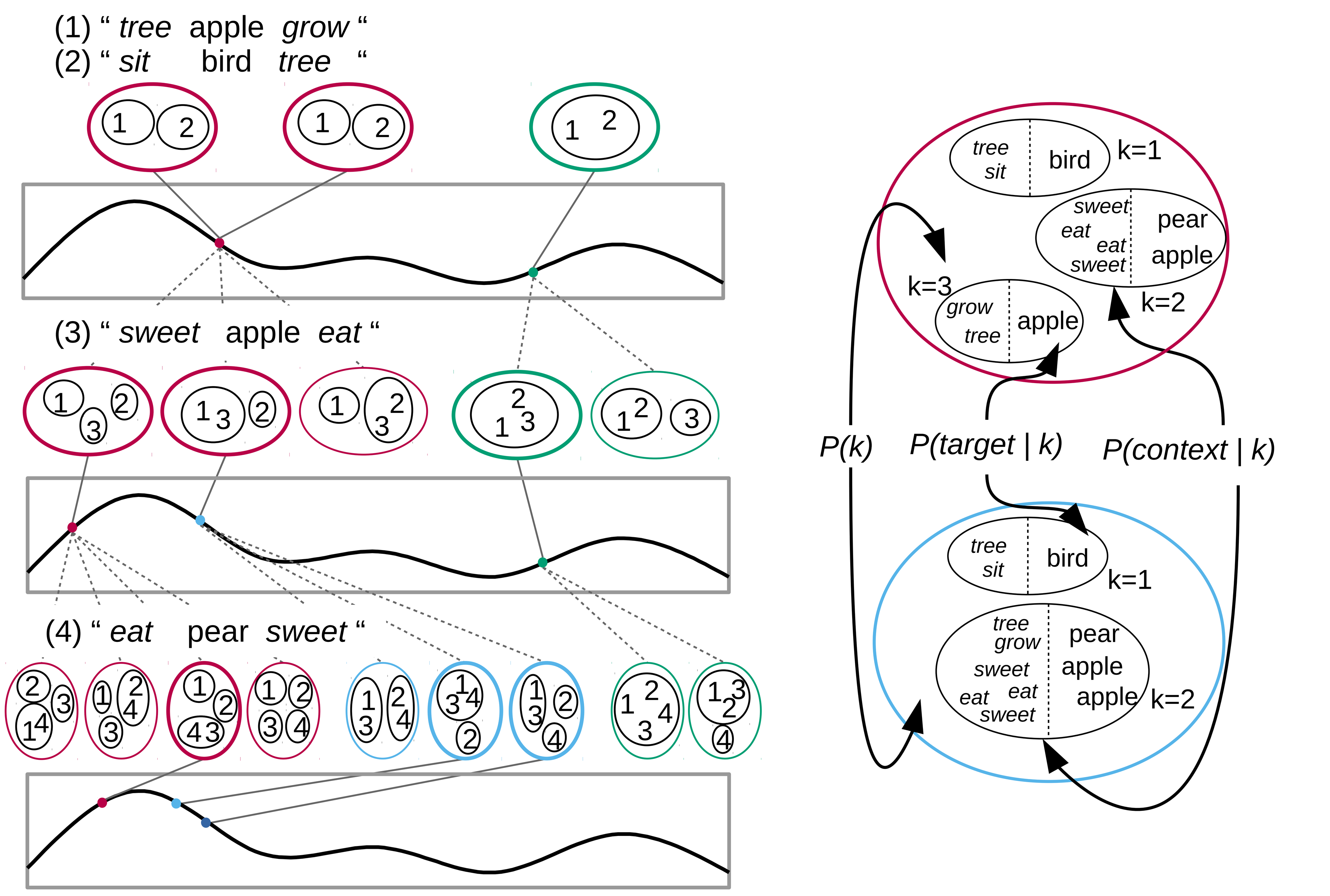
Assumptions

- A word's category is predictable from its linguistic context (≈ features)
- One global category distribution (θ), and separate target word (ϕ_k) and context word distributions (ψ_k) for each category k
- Concepts of a category co-occur with the same features but not necessarily with each other

Formally

- Multinomial category and word distributions, parameterized with conjugate dirichlet priors (efficient inference)
- Upper bound on the number of topics K is pre-specified as a number exceeding the number present in the data

Intuitive Example



Incremental Category learning using Particle Filtering

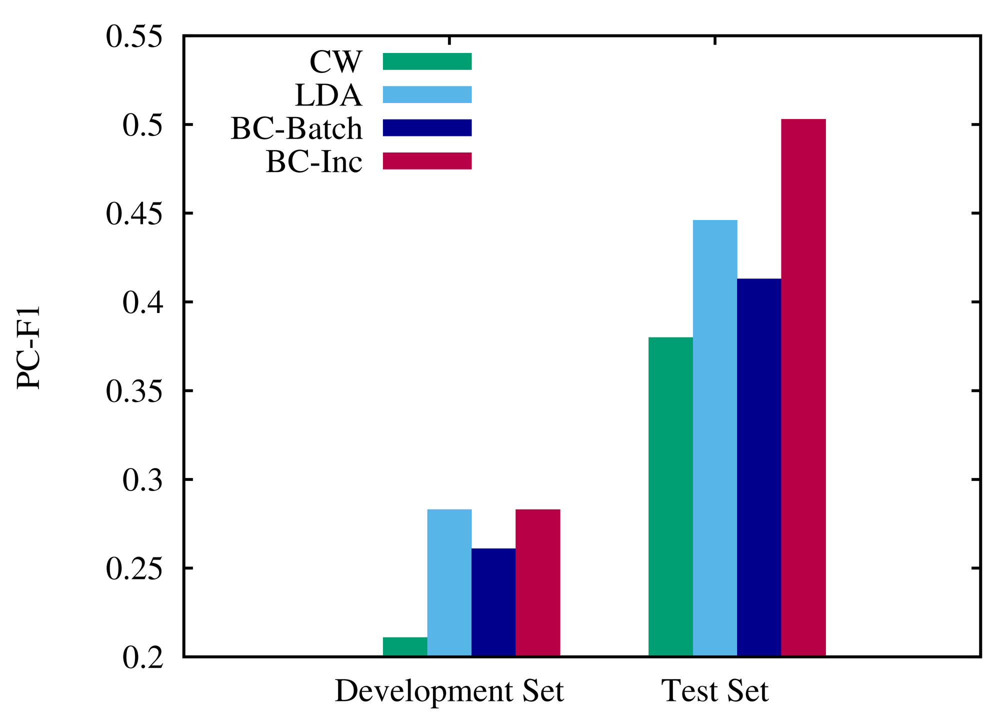
- Sequential Monte Carlo approximation of the true posterior distribution over categorizations $P(W, Z, \theta, \phi, \psi; \alpha, \beta, \gamma)$
- Propagate a set of categorization hypotheses (particles) through time
- Integrate each observation individually into each hypothesis and sample new particles from the distribution over all possible integrations

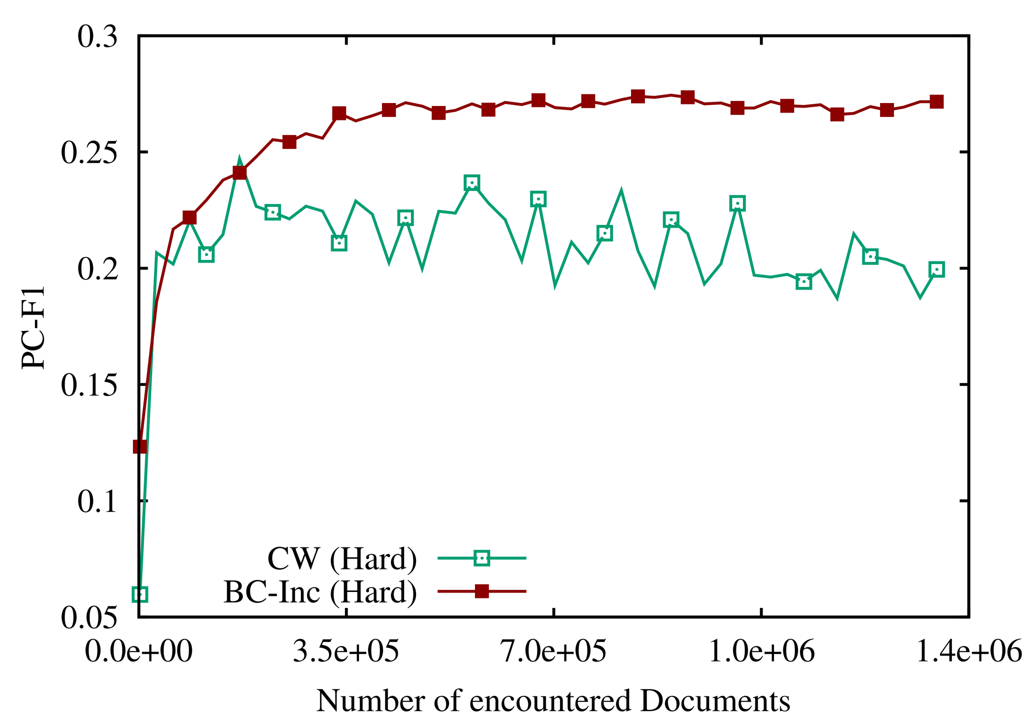
for particle p **do** ▷ Initialization
  Initialize randomly or from $z_p^0 \sim P_0(z)$
for observation t **do**
  for particle n **do** ▷ Sampling/Prediction
    $P_n(z_n^t|\mathbf{y}^t) \sim P(z_n^t|z_n^{t-1}, \alpha)P(y^t|z_n^t, \mathbf{y}^{t-1}, \beta, \gamma)$
  $\mathbf{z}^t \propto Mult(\{P_n(z_n^t)\}_{i=1}^N)$ ▷ Resampling

Experimental Setup

System Comparison	(CW) Graph-based model (Fountain et al (2011)), (LDA) Vanilla Topic Model, (BC-Batch) Batch BayesCat, (BC-Inc) Incremental BayesCat
Gold Standard	Categories based on human created Feature Norms (McRae et al (2005), Fountain et al (2010)); 70% development split, 30% test split
Data Set	British National Corpus (1.3M inputs ; 5 word context window)
Metric	(PC-F1) Purity/collocation F1

Experimental Results





Conclusions

- Incremental and simultaneous learning of categories and their features within one statistically sound framework.
- Outperform standard topic model and incremental clustering-based model

Future Work

- Learning abstract categories
- Learning hierarchical category structure (taxonomies)

Example Output

WEAPONS	
Concepts	shotgun, pistol, knife, crowbar, gun, sledgehammer, baton, bullet, motorcycle , van, ambulance
Features	injure, ira, jail, yesterday, arrest, stolen, fire, officer, gun, police, victim, hospital, steal , crash, murder

INSTRUMENTS	
Concepts	tuba, drum, harmonica, bagpipe, harp, violin, saxophone, rock, piano, flute, harpsichord banjo, guitar
Features	amp, orchestra, sound, electric, string, sing, song, drum, piano, condition, album, instrument