Filtering with Abstract Particles

Jacob Steinhardt, Percy Liang
Stanford University

Motivation

Goal Given an (un-normalized) target distribution $f^*(x)$, $p^*(x) = \frac{1}{Z}f^*(x)$, want to compute normalization constant Z.

- Extends naturally to marginals / conditionals, but focus on Z for concreteness.

Issue Often computationally intractable, so use some approximation $\hat{f}$ to f^*.

Illustration: Variational vs. Particle Methods

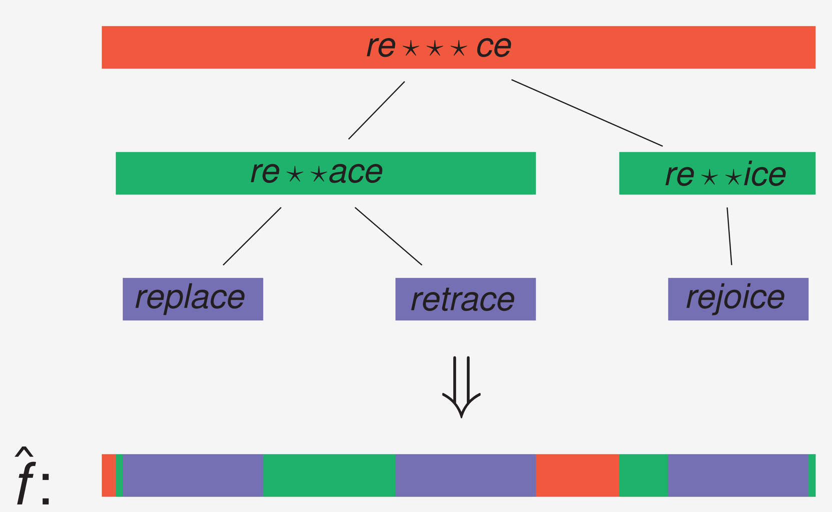
Goal: infer missing characters in `r e _ _ _ c e`

Particle	Actual	Variational
0.5 replace	0.33 replace	re [0.33 j, 0.33 p, 0.33 t] [0.33 l, 0.33 o, 0.33 r] [0.66 a, 0.33 i] ce
0.5 retrace	0.33 retrace	
	0.33 rejoice	
	0.01 ...	

Particles provide **precision** but lack **coverage**, while variational inference lacks precision.

Stitching Together Models

Question How to combine the different models?



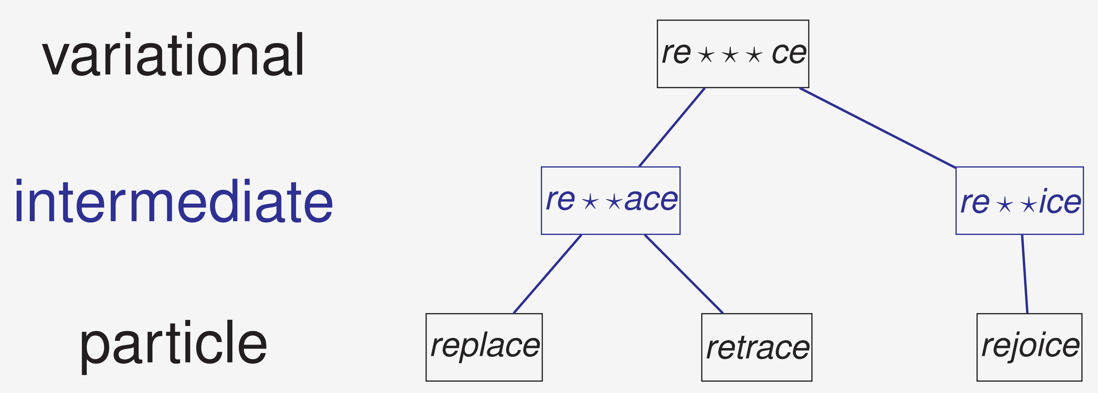
Answer Just use most precise model available at each point (relies on nested structure, i.e. the regions form a ***hierarchical decomposition***).

Contributions

- Introduce **abstract particles**, which interpolate between **particle** and **variational** inference.
- Identify **large, tractable** family of abstract particles using hierarchical decompositions, including models with **high tree-width**.
- Demonstrate **improved performance** over beam search and sequential Monte Carlo.

Our Proposal

Define approximations over intermediate **regions**.



Goal Stitch together approximations at multiple levels to simultaneously obtain precision (from lower levels) and coverage (from higher levels).

Inference

Can compute $\sum_{x \in X} \hat{f}(x)$ as long as we can compute sums over each region. Proof:

A Family of Approximations

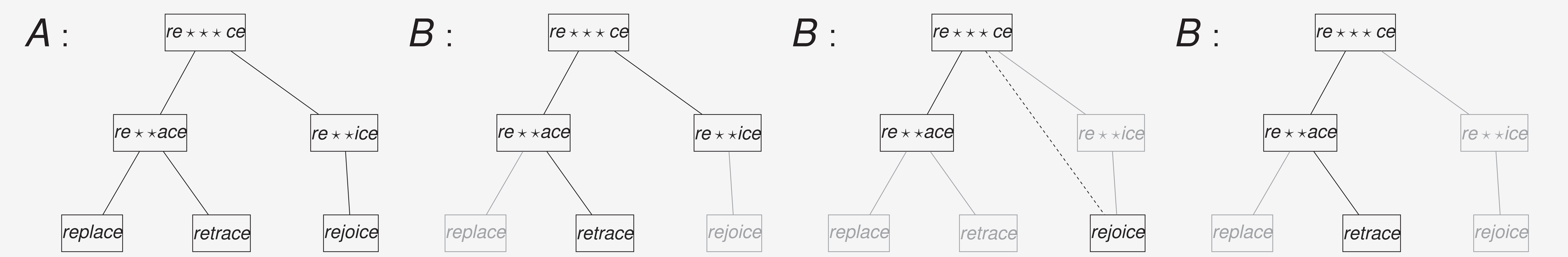
A hierarchical decomposition A leads to an approximation $\hat{f}$ (see bottom left panel).

We would like to define a family of approximations and choose the best one.

Key idea Every subset B of a hierarchical decomposition A is itself a hierarchical decomposition.

- Can let A have large cardinality and search for a small cardinality subset B that yields a good approximation.

Example of A and several possible subsets B:

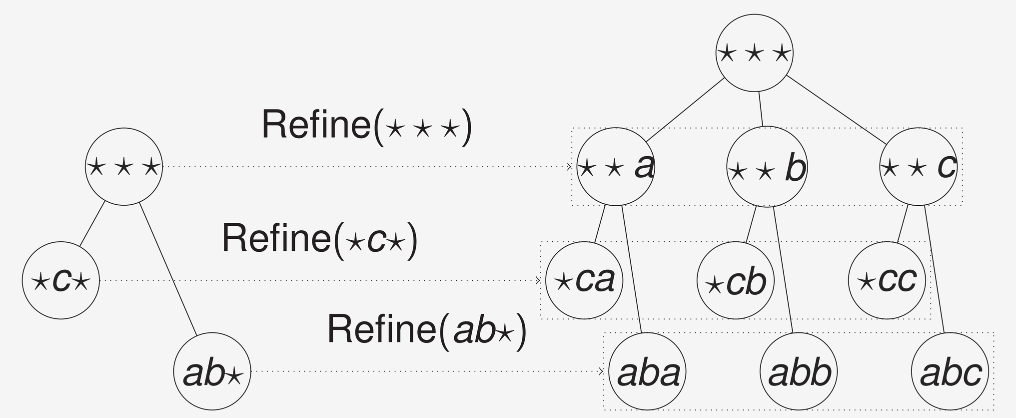


Search Strategy

Suppose that A has size 1000 and we want a subset of size 100: $\binom{1000}{100}$ possibilities; far too many!

Solution "Abstract beam search." Iteratively *refine* and *prune* a candidate decomposition.

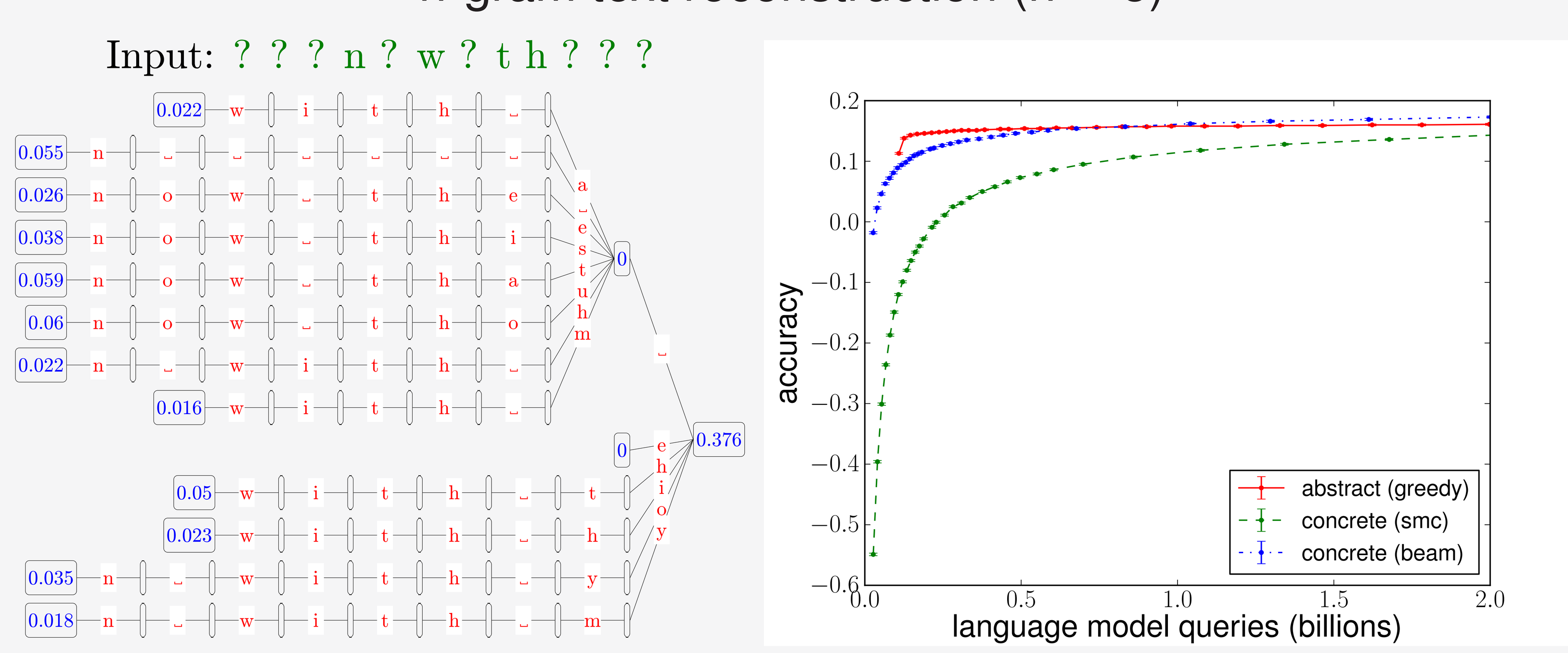
- **Refine**: split each region into smaller regions (to gain precision).
- **Prune**: greedily keep a small number of regions that yield a good approximation (so we can refine again).



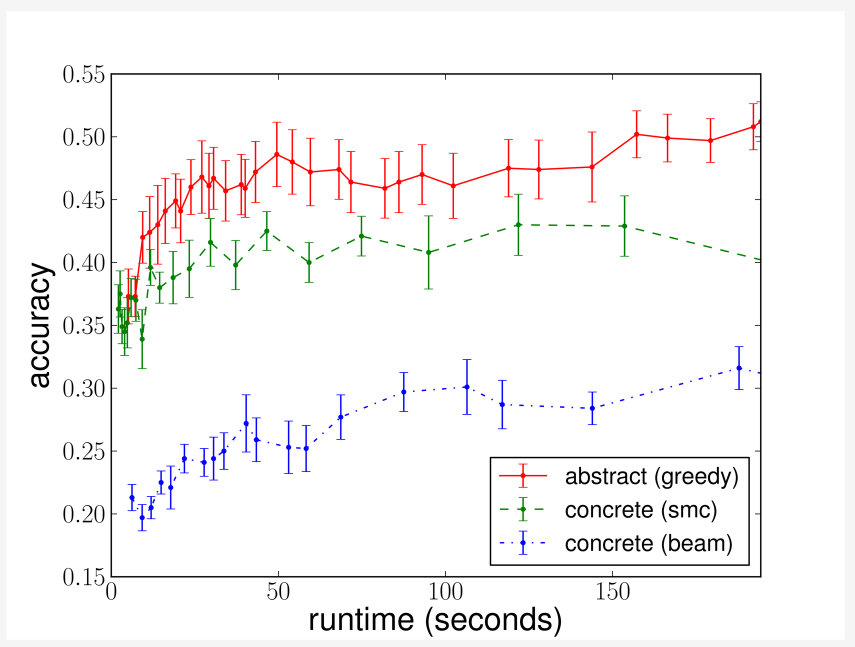
Applies naturally to filtering tasks (refine on current state to go to next time step, prune to save resources).

Experiments

n-gram text reconstruction ($n = 8$)



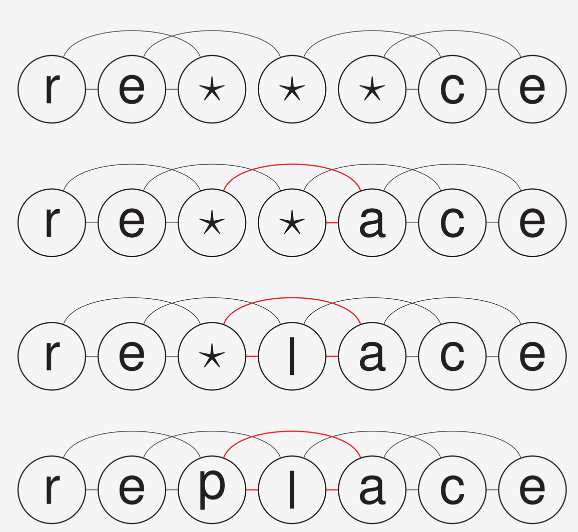
Factorial HMM (100 states, 15 factors)



Generalizing the Construction

Let X be some space. Suppose we have a hierarchical decomposition $A \subseteq 2^X$ together with an approximation $\hat{f}_a$ to f^* defined on each region $a \in A$.

- If $a = \{x_0\}$ is a singleton set (bottom), can have $\hat{f}_a(x_0) = f^*(x_0)$.
- If $a = X$ (top), will need to drop most of the dependencies.
- For intermediate values of a (for instance, fixing the values of certain variables) can keep some subset of the dependencies.



Set $\hat{f}(x) \stackrel{\text{def}}{=} \hat{f}_a(x)$, where a is the smallest region containing x.
Can think of each region $a \in A$ as an *abstract particle*.

Discussion

- Abstract particles combine the advantages of variational and particle inference.
- Hierarchical decompositions provide a framework for reasoning about the optimal representation for approximate inference.
- Related work: split variational inference (Bouchard & Zoeter, 2009).
- Also: coarse-to-fine inference (Petrov et al., 2006; Weiss & Taskar, 2010; many others).

The first author was supported by the Hertz foundation.