

## どんな研究？

- 確率モデルに基づいて、膨大なデータから重要で特徴的な性質を発見し理解することを可能にする技術
- 情報が生成された仕組みをモデル化し、隠れた構造を自動的に学習する技術

## もたらされる変革

- 音声データからいつ誰がしゃべったかを推定
- 文書データから隠れたグループを抽出
- あらゆる言語を辞書なしで解析
- 価値のあるキーワードを自動的に認識

## 展示紹介

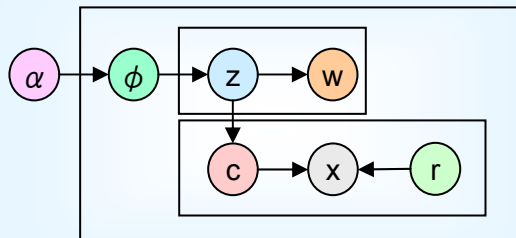
- 「会話」「文書」「文」「キーワード」が生成される過程をモデル化
- データに隠された重要で特徴的な性質「話者」「著者・文書グループ」「単語」「内容との関連性」を発見

- ① 話題の時間変化を追跡する
- ② 誰が私と似てる？どの文書がこれに似てる？
- ③ 隠れた単語を発見する：源氏物語からblogまで
- ④ 多様なキーワードから価値あるものを抽出する



連絡先： 岩田具治  
協創情報研究部  
知能創発環境研究グループ

## データの生成過程を記述するモデル ⊕



データがないときのモデルの分布

データから学習したモデルの分布

データの生成過程

$$P(\Theta | X) \propto P(\Theta)P(X | \Theta)$$

データからモデルを学習

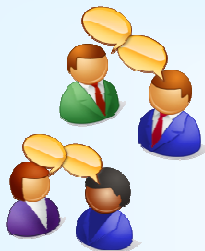
ベイズの定理

モデルからデータを生成

学習

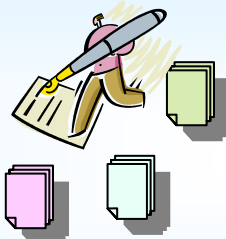
ベイズの定理に基づいてデータを生み出した原理（モデル）を学習

## 実際に観測されるデータ X



会話データ

いつ誰が話したかを特定



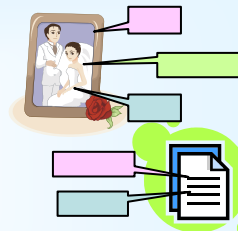
著者・文書データ

隠れたグループを抽出

その中から不要な情報を捨て、有益な知識を取り出すことは、ますます難しく...

言語データ

隠れた単語を発見



キーワードデータ

価値のあるキーワードを認識