

データを生み出す隠れた原理を発見する

誰が私と似てる？どの文書がこれに似てる？

－著者の興味と文書内容の関係に着目した潜在変数モデル－

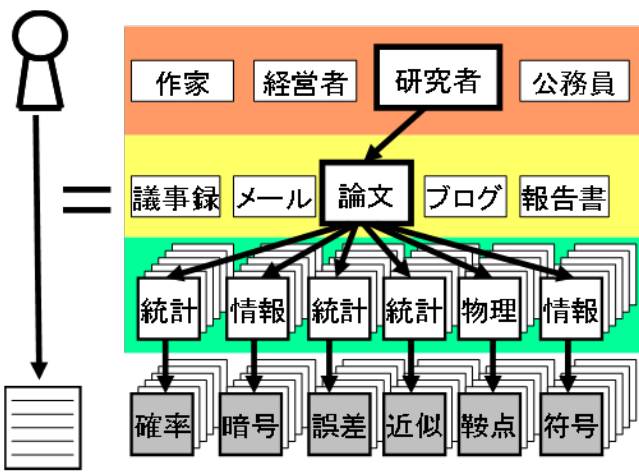
どんな研究？

「人が文書を書く」といった人→物の関係データを、より低次元の確率的な構造でモデル化を行う研究です。文書データに適用することで、「著者の興味→文書の内容」という文書生成のプロセスをモデル化し、著者間及び文書間それぞれの類似性を評価できます。

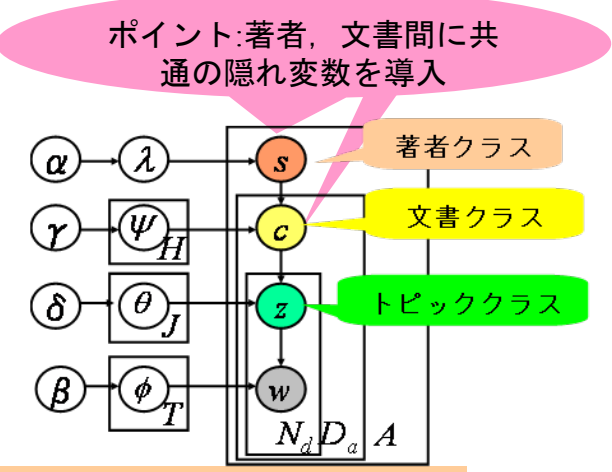
もたらされる変革

興味と内容を同時に用いた検索の実現
・世界中から自分と興味の似たグループ、そのグループで関心の高い文書を検索することができるようになる
・今まで知らなかった、興味に合ったWebページを見つけたりすることができるようになる

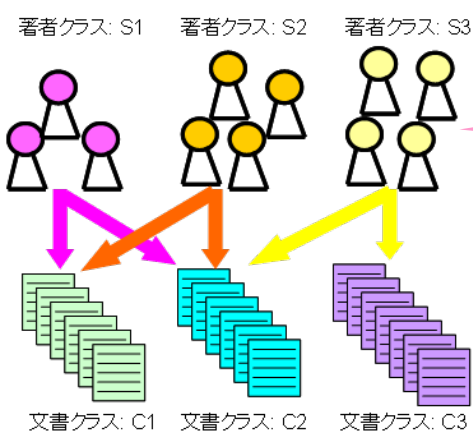
①ゴール：文書生成のプロセスのモデル化



②アイデア：類似した内容の文書はトピック生成の構造も似ている



③アウトプット：似ている著者及び文書それぞれをグループ化



著者を作成した文書の内容の類似性によりグループ化

文書をトピックの類似性によりグループ化

関連文献

川前徳章 山田武士: 著者の興味と文書の内容の関係に着目した潜在変数モデル, 電子情報通信学会 人工知能と知識処理研究会 (AI), 2009

連絡先: 川前徳章 (Noriaki Kawamae)
創情報研究部 知能創発環境研究グループ

