

# P5-26 ラベルノイズに頑健なトピックモデル

岩田具治 山田武士 上田修功 (NTT CS研)

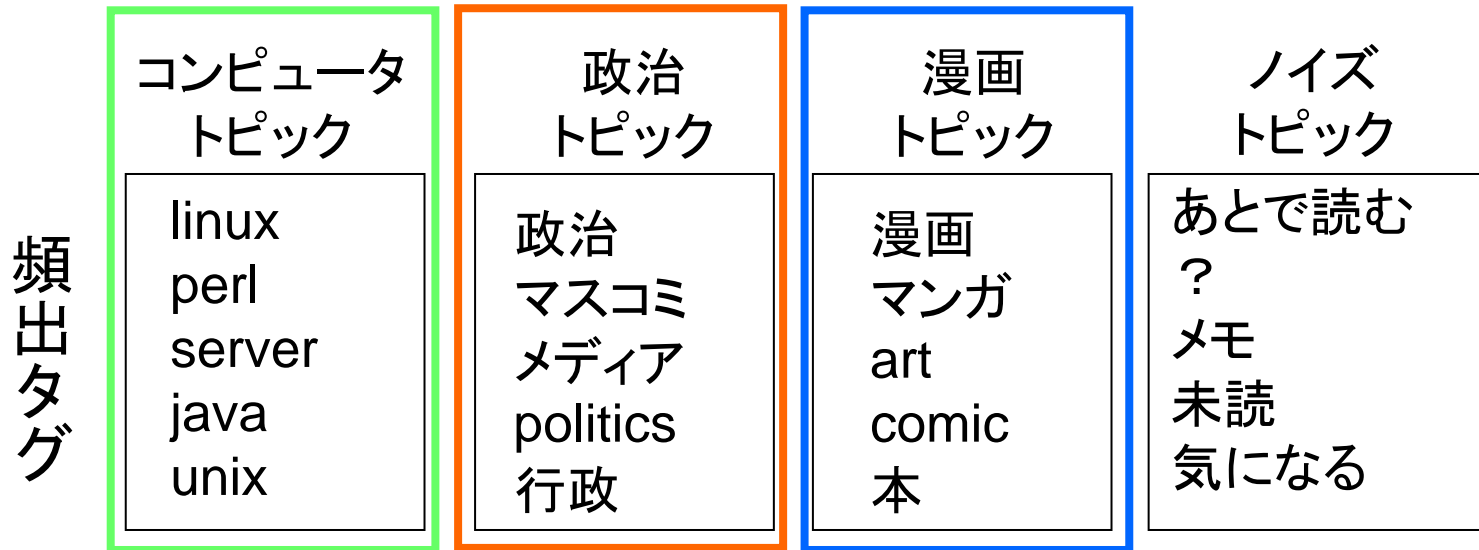
- ソーシャルアノテーション

- Hatena, Flickr, CiteULike, youtube, Last.fm
- ユーザが任意にラベルを付与できるため、内容に関連しないラベルが含まれる
- あとで読む, これはすごい, nikon

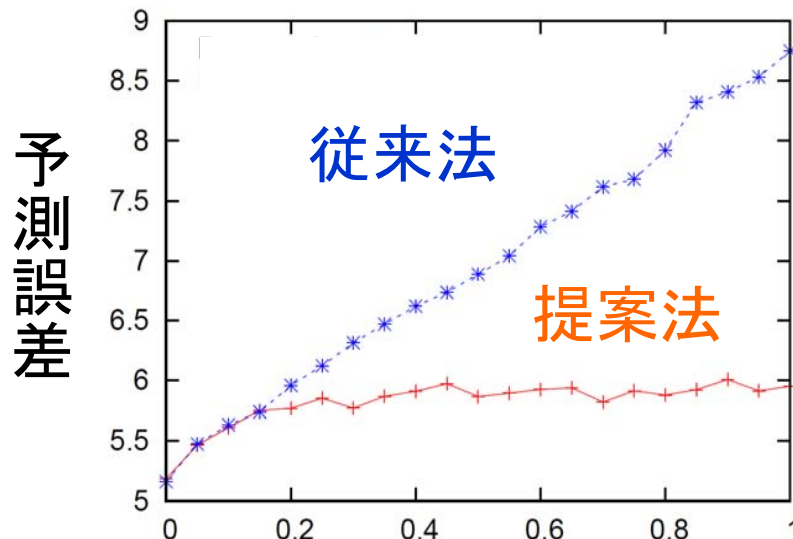
- ラベルノイズを含むデータに対しても頑健なトピックモデルを提案

- 推定: 確率的EMアルゴリズム

# ソーシャルブックマークからのトピック抽出



## タグ予測誤差比較



従来法に比べ  
ノイズが多い場合でも  
誤差が低いことを確認

ノイズ割合