

P010

# 品質の異なる二つのデータ集合間 の転移学習の解析

An Analysis of Transfer Learning between  
a Pair of Datasets with Different Qualities

赤穂昭太郎

神嶋敏弘

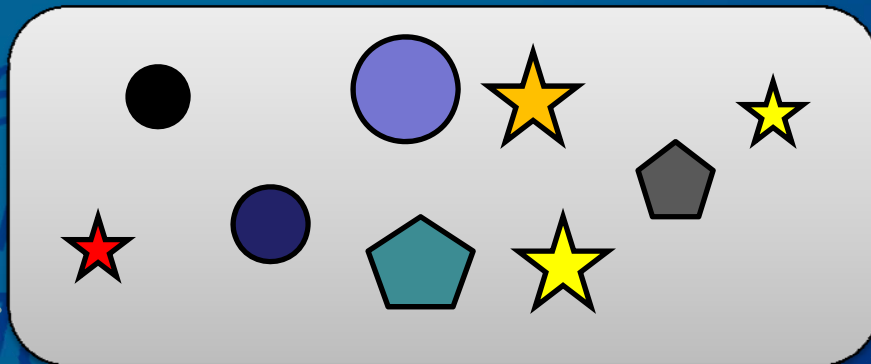
Shotaro Akaho

Toshihiro Kamishima

産総研 s.akaho@aist.go.jp, mail@kamishima.net



高品質・高価・少数サンプル



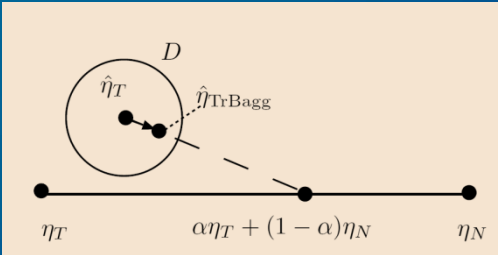
低品質・安価・大量サンプル  
高品質データが(少し)混ざっている

両方のサンプルを使って  
高精度の推定をしたい

# 単純なモデルでの解析

- バギングに基づいたアルゴリズム(神嶋)
- タグ付け問題で良好な結果(Kamishima et al. ICDM2009など)

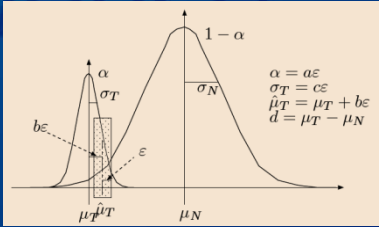
どんなときにうまくいく・いかないのか？



- **結果1:**  
単純な指数分布族の推定問題ではうまくいかない  
(ゴミに引っ張られる)



- **改良アルゴリズムの提案**
- **結果2:**  
正規分布の位置パラメータの推定  
改良アルゴリズムが緩い条件で真の値との誤差を小さくできる



$$\gamma \simeq 2 \frac{\phi(d, \sigma_N^2)}{a} + 2\phi(1, c^2)$$