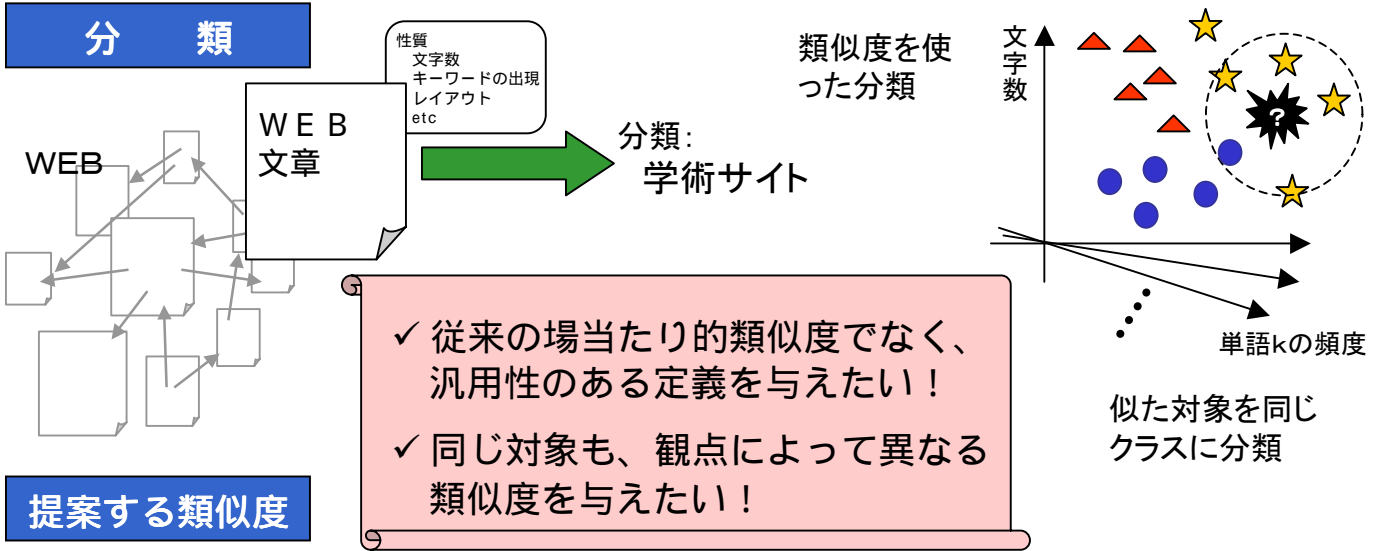


データ対の共通的性質を利用した統計的データ分類法

山田 泰大、犬塚 信博、世木 博久、和田 幸一
 キーワード：データの分類、データマイニング、WEB

データ間の類似性を使って、データを分類することができます。2つのデータを結びつける共通的性質を加味して確率的に類似性を決める技術を開発しました。WEB（ホームページ）の分類等への応用が期待できます。



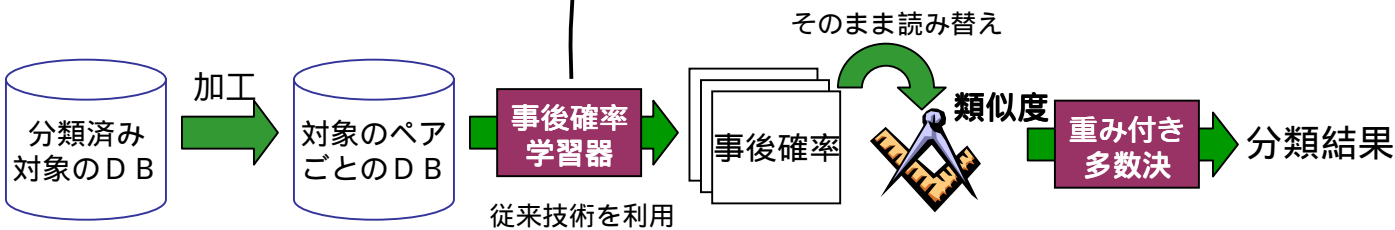
提案する類似度

2つの対象の類似度
 $\text{sim}(x, y)$

= 定義 2つの対象が同じ分類になる確率
 $\text{Pr}(\text{class}(x)=\text{class}(y) | x, y)$

(2つの対象の諸性質の観測に対して、分類クラスが同じになることの事後確率)

分類の流れ



本技術の特徴

- ✓ 2つの対象の共通的性質を積極的に利用できる。
- ✓ 分類の観点に応じて異なる類似度が導かれる。
- ✓ 類似度の意味が明確（類似した対象の投票による分類をした場合、最大事後確率分類と一致）

この技術を使うために必要なもの
 > 分類済み対象のデータベース
 (対象の性質と、分類クラス)
 データ間の共通的性質がある場合に効果的

応用分野

WEBサイト、学術論文、オークション等の分類