

# 集合知プログラミング

## 第3章 グループを見つけ出す

06T4007A 伊藤輝将

# はじめに

- 第2章

- 近い関係のものたちを探し出す方法について

- 例:映画について自分と同じような好みを持っている人々を探し出す

- 第3章

- データクラスタリングについて

# データクラスタリング

- ・お互いに関連しているもの同士、人々の集団、アイデアのグループなどを発見し、可視化するための手法の一つ
- ・教師なし学習の一例

# 教師あり学習

- ・目標や正解が入力されている場合の学習のこと

例：ニューラルネットワーク、決定木、  
サポートベクトルマシン、ベイジアンフィルタ

# 教師なし学習

- ・目標や正解とすべき情報が入力されない場合の学習のこと

- ・模範解答によって訓練されるものではない

例: クラスタリング、非負値行列因子分解、  
自己組織化マップ

→あとの二つについては第10章で取り上げる

# ブロガーを分類する

- ・それぞれのブログのフィードに特定の単語たちが現れた回数を基にクラスタリングする



同じサブジェクトについて頻繁に書いたり、同じようなスタイルで書いているブログのグループを決定できる



膨大な数のインターネット上のブログを検索、分類したり、探し出すのに非常に役立つ

# 階層的クラスタリング

- ・最も似ている二つのグループをまとめることを繰り返すことによって、グループの階層を作りあげること
- ・グループが1つしか存在しなくなるまで繰り返す

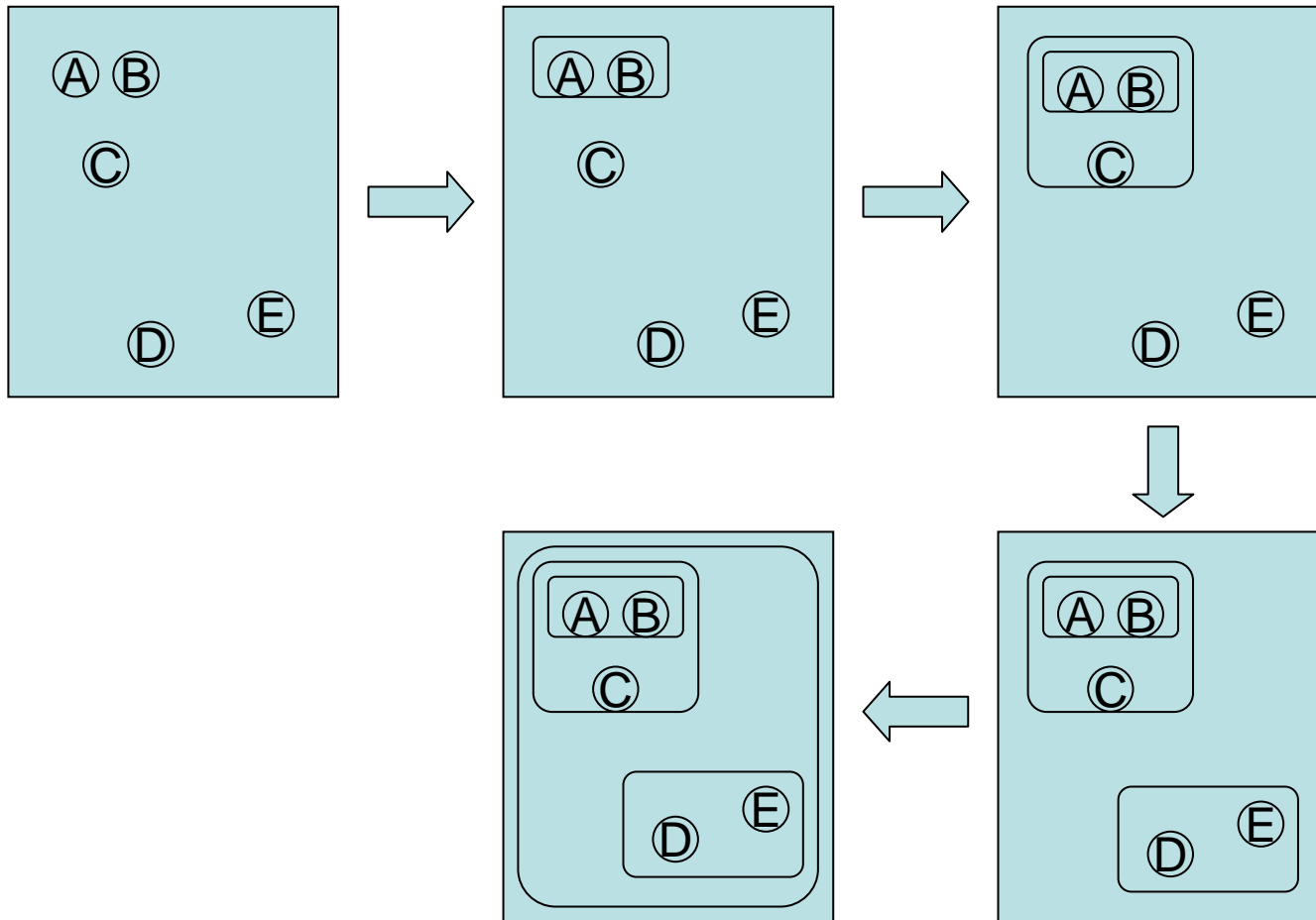
# 階層的クラスタリング(2)

- ・階層的クラスタリング終了後、通常は、デンドログラムと呼ばれるグラフの形式で結果を確認する

- デンドログラム

- ノードをピラミッドの形で並べて表示するグラフ

# 階層的クラスタリングの動作



# 階層的クラスタリングのアルゴリズム

- ・最初のクラスタは元々のアイテム



すべての組をループし最も距離の近い組を探す



二つのクラスタの平均を計算する



新たなクラスタを作る