

交友関係データと脳波データからの 関連性自動形成

日本大学文理学部
情報システム解析学科

谷聖一研究室
小林 彩

背景・研究動機

- コンピュータゲームやコンピュータ操作等が人間の脳に与える影響について議論されている。
- 悪影響があるとし、
影響を受けた脳をゲーム脳と定義 (森昭雄 2002年)

独自開発の簡易脳波計で、ゲーム中の脳波を測定する実験を行う事によって、「コンピュータゲーム、携帯電話のメール入力、パソコン操作が人間の脳に悪影響を与える」と主張。

「長時間一緒にゲームをしている友達同士は
ひとしく脳波に悪影響が出る」という仮説が存在。

背景・研究動機

- データマイニングの代表的な手法を用いて、交友関係データと脳波データの関係性を抽出。
※データマイニングについてはこの後説明。
- 本研究では、小学生を対象とした調査データを用いる。

小学生は比較的テレビゲームで遊ぶ時間が長く、友達と一緒にゲームをして遊ぶ機会も多い。

脳波測定の結果と友達関係を照らし合わせて検討すれば、何らかの知見が得られるのではないかという考えに基づく。

「小学生の交遊関係は同じクラス、同じ学年でほぼ完結しているため、関係図を作るのが容易である」

データマイニングとは

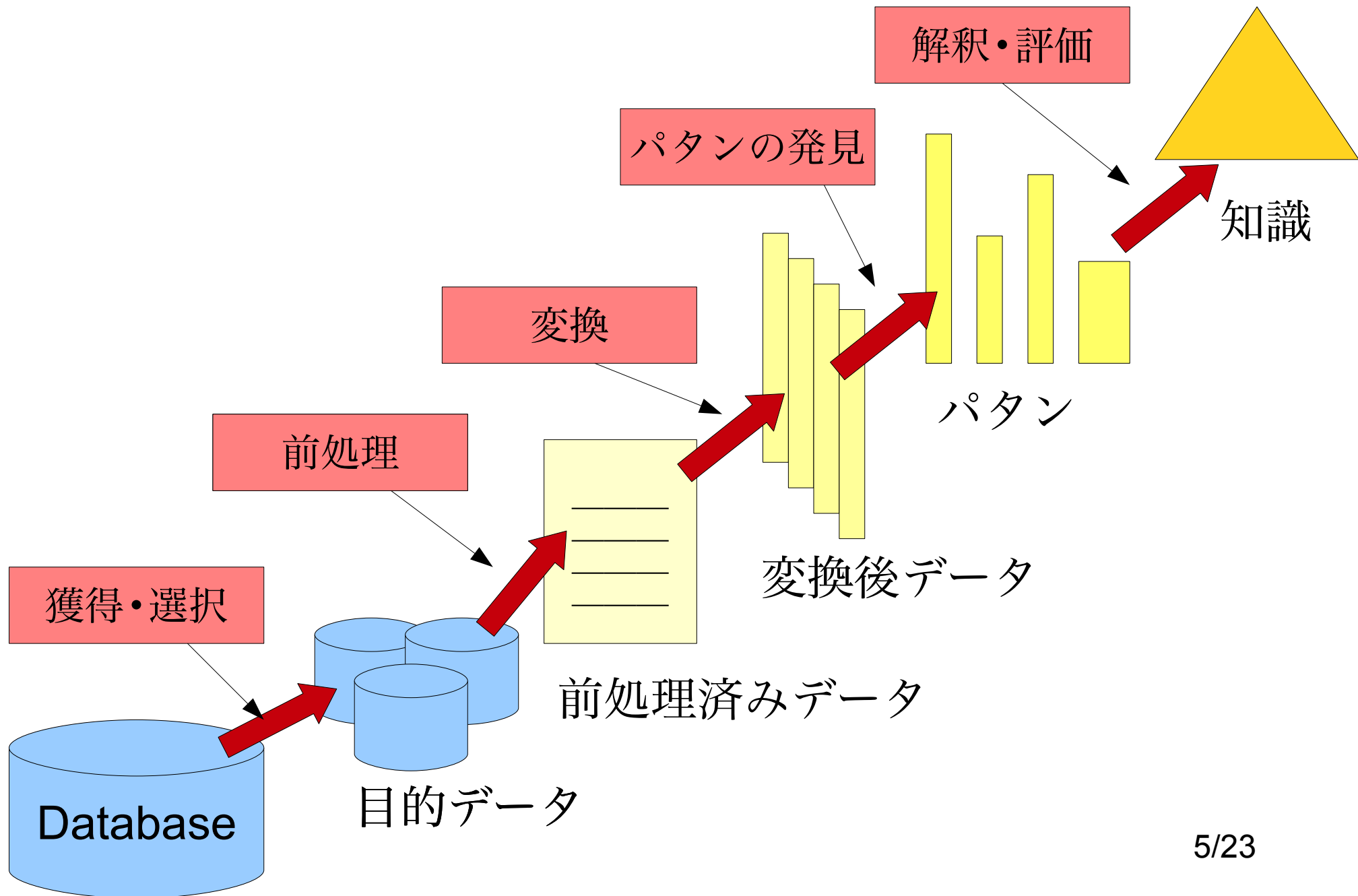
- Data Mining 略して DM
- 多量のデータから
有用な知識を発掘する技術の総称.

例 : 顧客管理情報, 売り上げデータ...

手法 : 頻出パターン抽出, クラス分類,
回帰分析, クラスタリング

- knowledge-discovery in databases
(データベースからの知識発見)
KDD と呼ばれる.

データマイニングのプロセス



データマイニングの手法

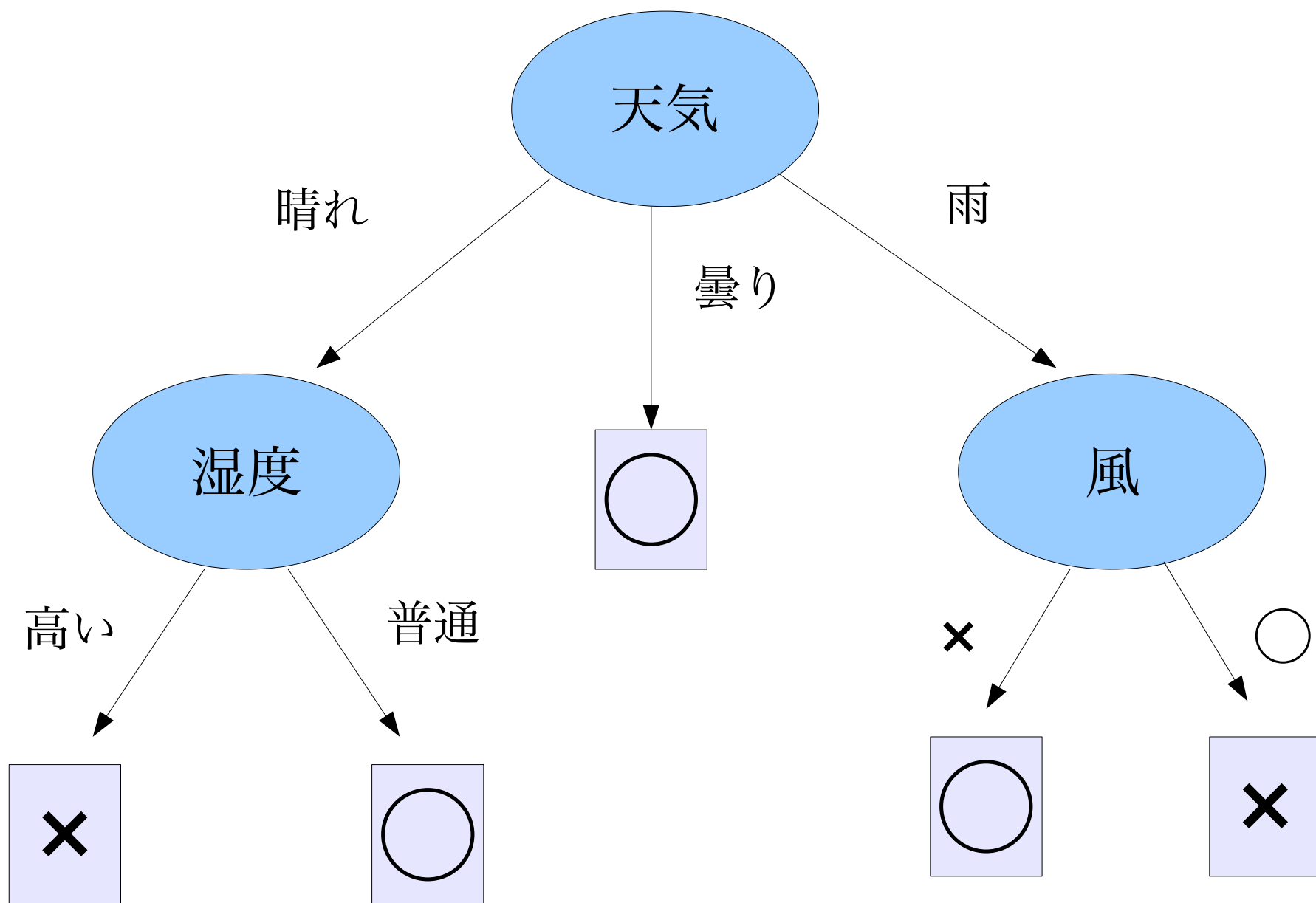
- クラス分類
- クラスタリング
- 相関ルール抽出

分類

決定木学習（データから決定木をつくる）

例：ゴルフが出来るかどうか

天気	温度	湿度	風	プレイ
晴れ	暑い	高い	×	×
晴れ	暑い	高い	○	×
曇り	暑い	高い	×	○
雨	暖かい	高い	×	○
雨	涼しい	低い	×	○
雨	涼しい	低い	○	×
曇り	涼しい	低い	○	○
晴れ	暖かい	高い	×	×
晴れ	涼しい	低い	×	○
雨	暖かい	低い	×	○
晴れ	暖かい	低い	○	○
曇り	暖かい	高い	○	○
曇り	暑い	低い	×	○
雨	暖かい	高い	○	×



クラスタリング

データの集合をクラスタと呼ぶグループに分ける.

同じクラスタのデータならば互いに似ていて
違うクラスタならば互いに似ていない.

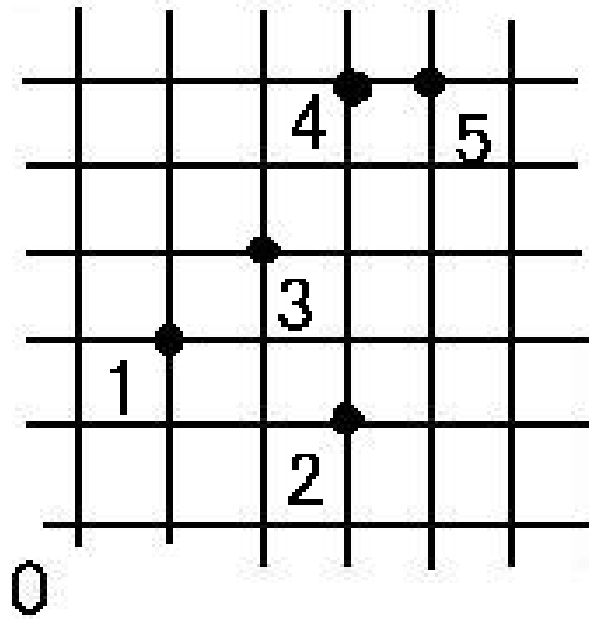
- 最遠点アルゴリズム,
- 最短距離法のアルゴリズム

最遠点アルゴリズム

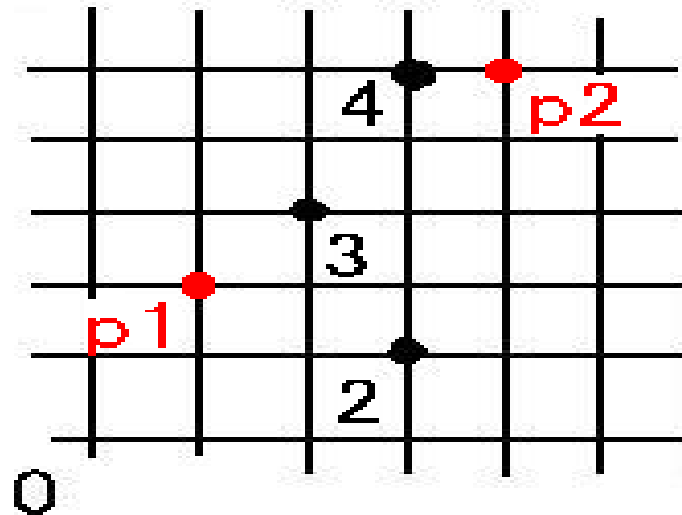
- 1, 任意の 1 点 $p1$ を S からとり, $P = \{ p1 \}$ とする.
- 2, S のうちで P からの距離のもっとも遠い点を計算し, P に加える.

初期 $P = \{ 1 \}$
 $P = \{ 1, 5 \}$

$p1 = 1$, $p2 = 5$ とする.



$S =$ 上記の 5 点の集合



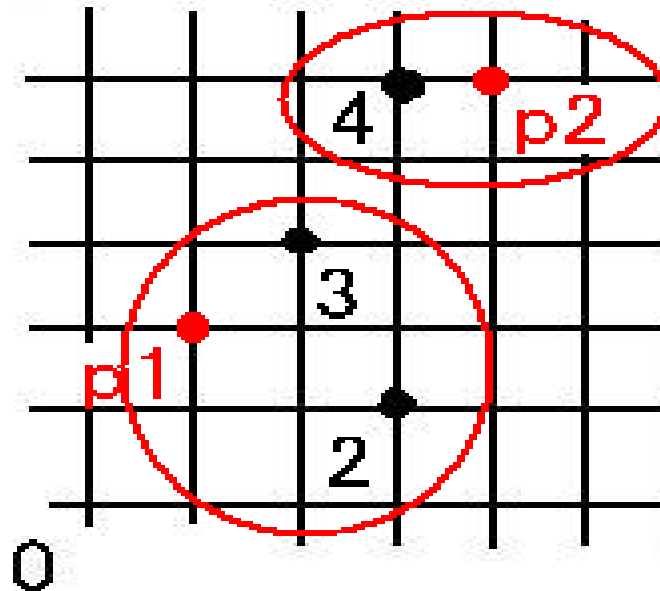
最遠点アルゴリズム

3, Pのサイズが k 未満なら 2 に戻る. (k=2) Kはアルゴリズムの実行前に決定して固定.

4, Pの点に対応した k 個のクラスタを作る.
このとき S の各点は最も近い P の点に対応したクラスタに分類される.

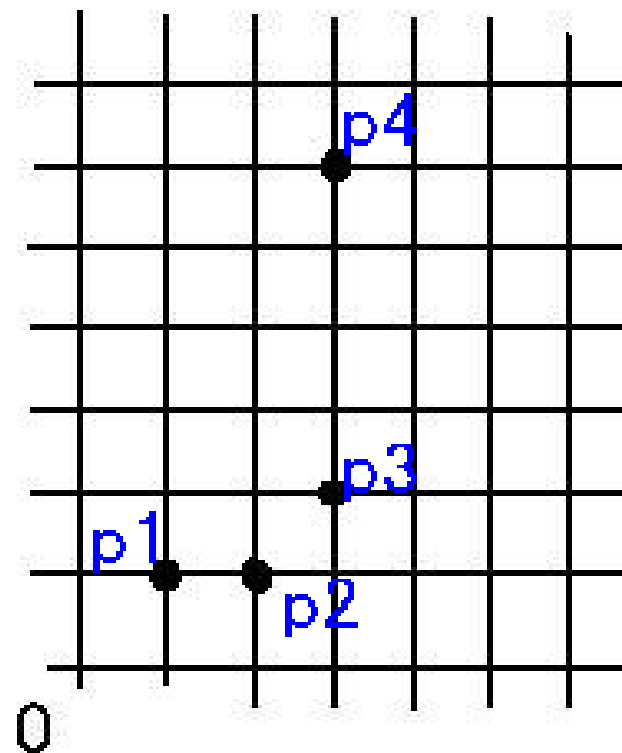
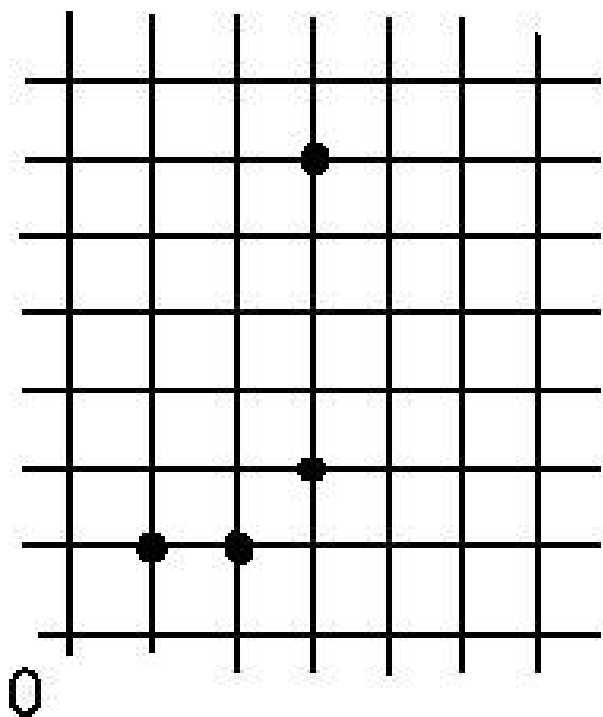
$$\begin{aligned} d(p1,2) &= \sqrt{(1-3)^2 + (2-1)^2} = \sqrt{(-2)^2 + 1^2} = \sqrt{5} \\ d(p2,2) &= \sqrt{(4-3)^2 + (6-1)^2} = \sqrt{1^2 + 5^2} = \sqrt{26} \end{aligned}$$

よって, { 1,2,3 }, { 4,5 }



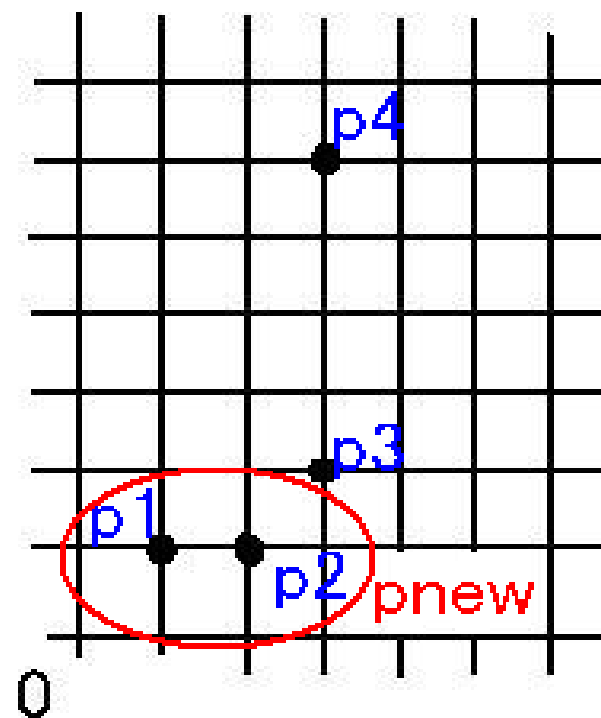
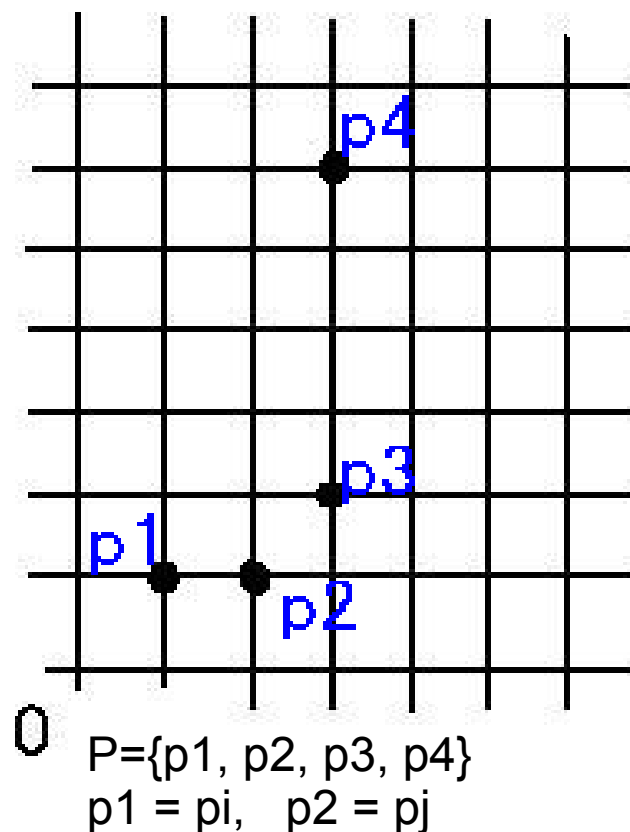
最短距離法のアルゴリズム

入力：各データを単一要素のクラスタとし、
クラスタ集合 $P = \{p_1, \dots, p_n\}$ とする。



最短距離法のアルゴリズム

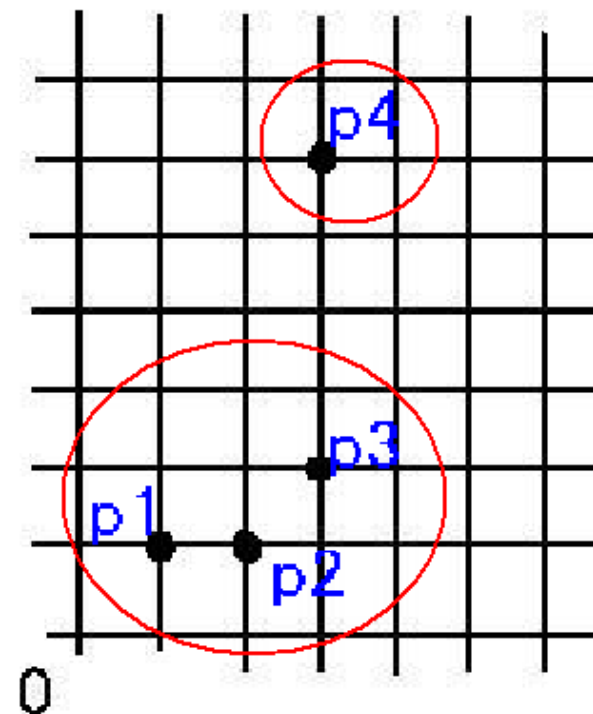
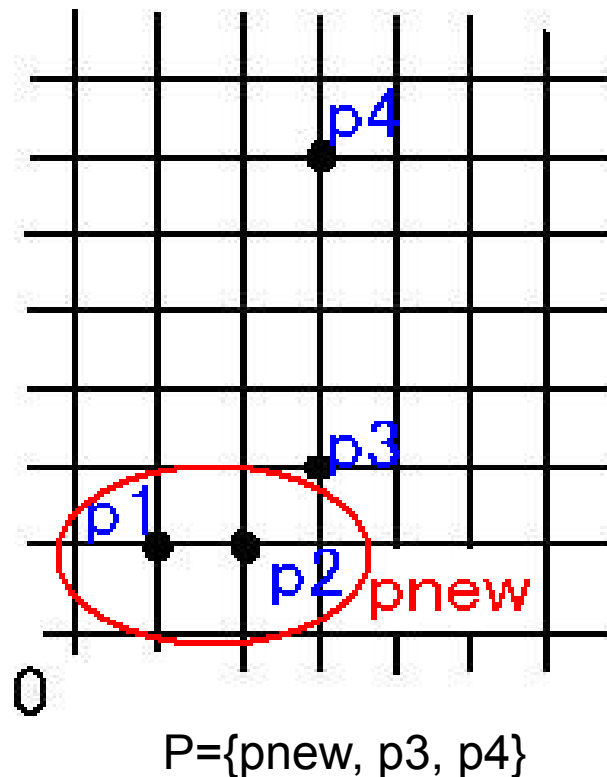
- 1, P の各クラスタ間の距離を求め,
最も距離の短いクラスタのペア p_i, p_j を選択.
- 2, p_i, p_j が同じクラスタに属していなければ,
 p_i と p_j を併合し, クラスタ p_{new} を作成.



最短距離法のアルゴリズム

3, P に p_{new} を追加し, p_i , p_j を削除.

クラスタ数が設定した数になっていなければ1に戻る.



相関ルール抽出

$A \rightarrow B$ という相関ルールは
Aという事象が起こると、Bという事象もおこりやすいことを意味する。

バスケット分析

(POS, Eコマースの取引ログに含まれる購買履歴を利用する)

例：紙おむつを買う人はビールも同時に購入する人が多い。



紙おむつの近くにビールを配置。

本Cを買う人は、後に本Dを買う事が多い。



本Cの購入者に本Dを薦めるメールをする。

アプリアリアルゴリズム

大規模データから逆単調性を利用し、
効率よく枝刈りを実行して多頻度アイテム集合を抽出する。

「 P が多頻度アイテム集合であれば
その部分集合 Q は多頻度アイテム集合である」

「 Q が多頻度アイテムでなければ
 Q を含むような集合 P も多頻度アイテム集合でない」

例: $\{A, B\}$ が多頻度アイテム集合でなければ
 $\{A, B\}$ を含むいかなるアイテム集合(例: $\{A, B, C\}$)も
多頻度アイテム集合でない。

アプリアリアルゴリズム

$$\text{集合Sの支持度} = \frac{\text{Sの要素の中で出現するトランザクション件数の最小}}{\text{全トランザクション件数}}$$

支持度：相関ルールが全トランザクションデータでどの程度出現するかを表す割合。

トランザクション：売買取引。

あるアイテム集合の支持度が最低支持度より小さくなったとき、**逆単調性**を利用する。

そのアイテム集合を含むようなアイテム集合は、多頻度アイテム集合の候補とせず枝刈りをする。

購買トランザクションデータ

TID	購買アイテム
1	ミルク, パン, バター
2	ミルク, パン, ジャム
3	ミルク, マーガリン
4	パン, バター
5	ミルク, パン, バター, ジャム
6	マーガリン
7	ミルク, パン, ジャム, マーガリン
8	ジャム

最小支持度：3/8

F1は3回以上出現する
購買アイテム.

$F1 = \{ \text{ミルク, パン, バター, ジャム, マーガリン} \}$

次にF1を利用して、
大きさ2の
多頻度アイテム集合候補
C2を生成.

TID	購買アイテム
1	ミルク, パン, バター
2	ミルク, パン, ジャム
3	ミルク, マーガリン
4	パン, バター
5	ミルク, パン, バター, ジャム
6	マーガリン
7	ミルク, パン, ジャム, マーガリン
8	ジャム

$F1 = \{ \text{ミルク, パン, バター, ジャム, マーガリン} \}$

$C2 = \{ (\text{ミルク, パン})4, (\text{ミルク, バター})2, (\text{ミルク, ジャム})3, (\text{ミルク, マーガリン})2, (\text{パン, バター})3, (\text{パン, ジャム})3, (\text{パン, マーガリン})1, (\text{バター, ジャム})1, (\text{バター, マーガリン})0, (\text{ジャム, マーガリン})1 \}$

アイテム集合の末尾の数字：
データベース上(右の表)の
C2のアイテム集合の出現回数.

最小支持度以上となる(=3回以上出現)
多頻度アイテム集合F2は...

$F2 = \{ (\text{ミルク, パン}), (\text{ミルク, ジャム}), (\text{パン, バター}), (\text{パン, ジャム}) \}$

TID	購買アイテム
1	ミルク, パン, バター
2	ミルク, パン, ジャム
3	ミルク, マーガリン
4	パン, バター
5	ミルク, パン, バター, ジャム
6	マーガリン
7	ミルク, パン, ジャム, マーガリン
8	ジャム

$$F2 = \{ (\text{ミルク}, \text{パン}), \\ (\text{ミルク}, \text{ジャム}), \\ (\text{パン}, \text{バター}), \\ (\text{パン}, \text{ジャム}) \}$$

さらにF2を利用して,
大きさ3の
多頻度アイテム集合
C3を生成.

$$C3 = (\text{ミルク}, \text{パン}, \text{ジャム}), \\ (\text{ミルク}, \text{パン}, \text{バター}), \\ (\text{パン}, \text{ジャム}, \text{バター}) \\ \times (\text{ミルク}, \text{バター}, \text{ジャム})$$

TID	購買アイテム
1	ミルク, パン, バター
2	ミルク, パン, ジャム
3	ミルク, マーガリン
4	パン, バター
5	ミルク, パン, バター, ジャム
6	マーガリン
7	ミルク, パン, ジャム, マーガリン
8	ジャム

ここで**逆単調性**を利用して、
C3から1 アイテム削除した大きさ2の
アイテムセットすべてがF2の中に存在するかを調べる。

$$C3 = (\text{ミルク}, \text{パン}, \text{ジャム})3, \quad \begin{array}{l} (\text{ミルク}, \text{パン}), (\text{ミルク}, \text{ジャム}) \\ (\text{パン}, \text{ジャム}) \end{array}$$
$$(\text{ミルク}, \text{パン}, \text{バター}) \times, \quad \begin{array}{l} (\text{ミルク}, \text{パン}), (\text{ミルク}, \text{バター}) \\ (\text{パン}, \text{バター}) \end{array}$$

(パン, ジャム, バター) × (パン, ジャム), (パン, バター), (ジャム, バター)

F2 = (ミルク, パン), (ミルク, ジャム), (パン, バター), (パン, ジャム)

F3 = (ミルク, パン, ジヤム)

このあと,F3からC4は作成できないため,ここで終了.

今後の目標

- データマイニング

85人の小学生の脳波, 性別, 友人一覧,
コンピュータゲームのプレイ時間を元に
交遊関係をグラフ化.

- 可視化

交遊関係を見やすい様に描画.

