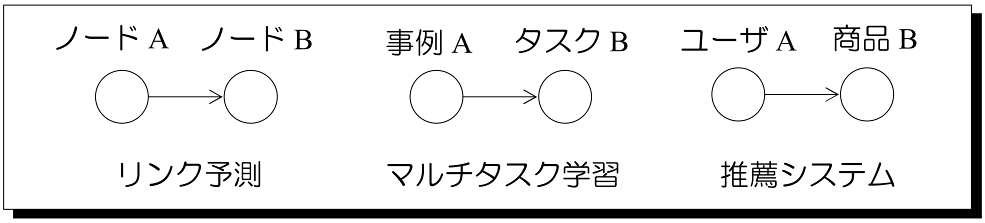
高速なペアワイズカーネルと汎化誤差解析について

鹿島 久嗣(IBM) / 小山 聡(京大)
山西 芳裕(Mines ParisTech) / 津田 宏治(MPI)

- 2つのインスタンスの間の関係（つながり／作用）を予測する問題を解くためには、インスタンスペア同士の類似度の定義が必要です
- そこで、私たちは、計算が簡単な、インスタンスペア同士の類似度「カルテシアン・カーネル」を提案します
- また、その性能について、カーネル（類似度）行列の固有値に基づいた考察をおこないます

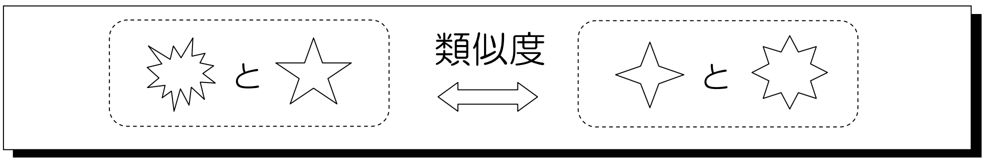
高速なペアワイズカーネルと汎化誤差解析について

我々は、ペアワイズ分類問題



をうまく解くための

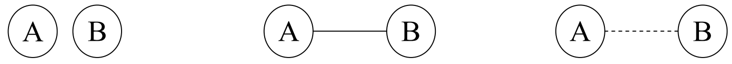
ペア同士の類似度の設計について考えます



ペアワイズ予測問題は、インスタンスペアの間の関係を予測する問題です

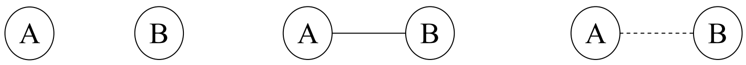
- たとえば、リンク予測は、ノード間の関係を予測する

2つのノード 間　に　リンクがある　のか　リンクが ない　のか



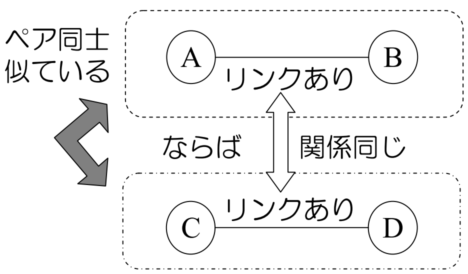
- たとえば、協調フィルタリングでは、ユーザーとアイテム間の関係を予測する

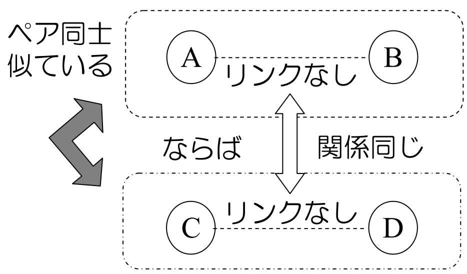
ユーザA が アイテムB を　買う　のか　買わない　のか



ペアワイズ予測を行うためには、ペアワイズ類似度が必要です

- 仮説「2つのペア同士が似ていれば、ペア間の関係も似ている」による推論を行いたい
- そのためには、ペア同士の類似度を適切に定義する必要がある

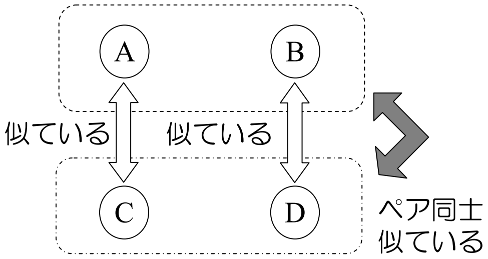




新しいペアワイズ類似度として「カルテシアン・カーネル」を提案します

- 既存のペアワイズ類似度

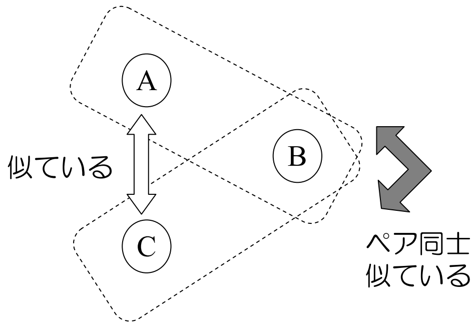
（クロネッカー・カーネル）



「ペアをまたいだペアが
両方とも似ていたら
ペア同士が似ている」とする

- 提案するペアワイズ類似度

（カルテシアン・カーネル）



「ペアをまたいだペアの
片方が共通で、
もう片方が似ていたら
ペア同士が似ている」とする

提案するペアワイズカーネルのカーネル行列は、インスタンス間カーネル行列のクロネッカー和で定義されます

- 既存のペアワイズ類似度

（クロネッカー・カーネル）

$$K \otimes K$$

  - インスタンス間カーネル行列のクロネッカー積
  - 重みつき完全グラフのクロネッカー積グラフとしても解釈できる
  - 密で巨大なペアワイズ・カーネル行列のため遅い

- 提案するペアワイズ類似度

（カルテシアン・カーネル）

$$K \oplus K$$

  - インスタンス間カーネル行列のクロネッカー和
  - 重みつき完全グラフのデカルト積グラフとしても解釈できる
  - 疎なペアワイズ・カーネル行列のため高速

2つのペアワイズカーネルの性能を理論的に比較してみます

- 訓練集合での経験リスクをもとに、2つのペアワイズカーネルの性能を仮説の期待リスクを評価します
- 全てのデータ点 $X=\{x_1, x_{2,\ldots,} x_M\}$ が既知の分類問題を仮定

（ペアワイズ分類では　$X = V^{(1)} \times V^{(2)}$）

- 仮説 h の期待リスク（予測誤りの確率）：

$$R(h) = \sum_{(x,y)\in Z} \delta(h(x)y \leq 0) P(x,y)$$

- サイズ $m<M$ の訓練集合 $\{(x_i, y_i) : i \in \{1,\ldots,m\}\}$ での経験マージンリスク：

$$R_s^\gamma(h) = \frac{1}{m}\sum_{i=1}^{m} \delta(h(x_i)y_i < \gamma)$$

期待リスクの上限は、カーネル行列の固有値の分布をもとに見積れます

[定理 (Kroon 2003)] $\lambda_1 \geq \lambda_2 \geq \Lambda \geq \lambda_M$ を全データ点から計算されるカーネル行列の固有値とする．仮説クラスを $\mathcal{F}(c)_B = \{\langle w, x\rangle + b : \|w\| \leq c, |b| \leq B\}$ としたとき，全ての $\gamma \in (8\Upsilon(n), 1]$ に対して同時に以下が成り立つ

$$P_{s \in Z^m}\left(\exists h \in \mathcal{H}(c)_B : \underbrace{R(h)}_{\text{期待リスク}} \geq \underbrace{R_s^\gamma(h)}_{\text{経験リスク}} + \sqrt{\underbrace{\frac{\left(n \ln 2 + \ln\left(\frac{\lceil c \rceil}{\theta\gamma}\right)\left\lceil\frac{8B}{\gamma}\right\rceil\right)}{2m}}_{\text{固有値から計算される量}}}\right) \leq \theta$$

ただし

$$\Upsilon(n) = \min_{j\in\{1,\ldots,n-1\}} 6 \cdot 2^{-\frac{(j-1)}{k(2^{j-1})}} (\lambda_1 \cdots \lambda_{k(2^{j-1})})^{\frac{1}{2k(2^{j-1})}} c(n,j)$$

$$k(l) = \min\left\{k \in \{1,\ldots,M\} : \lambda_{k-1} \leq \left(\frac{\lambda_1 \cdots \lambda_k}{l^2}\right)^{\frac{1}{k}}\right\}$$

$$c(n,j) = \min\left(1, 1.86\sqrt{\frac{\log_2\left(\frac{M}{n \; j} + 1\right)}{n-j}}\right)$$

ペアワイズカーネル行列の固有値は、インスタンス間のカーネル行列の固有値から簡単に計算できます

- カーネル法の汎化誤差を、カーネル行列の固有値の分布を用いて評価する方法が知られています
- ところが、ペアワイズカーネル行列のサイズは巨大であるため、直接固有値を計算することは大変です
  - たとえば、500ノードの場合でも、250,000×250,000 の行列になる
- 実は、ペアワイズカーネル行列の固有値は以下の定理を利用して、簡単に計算できます

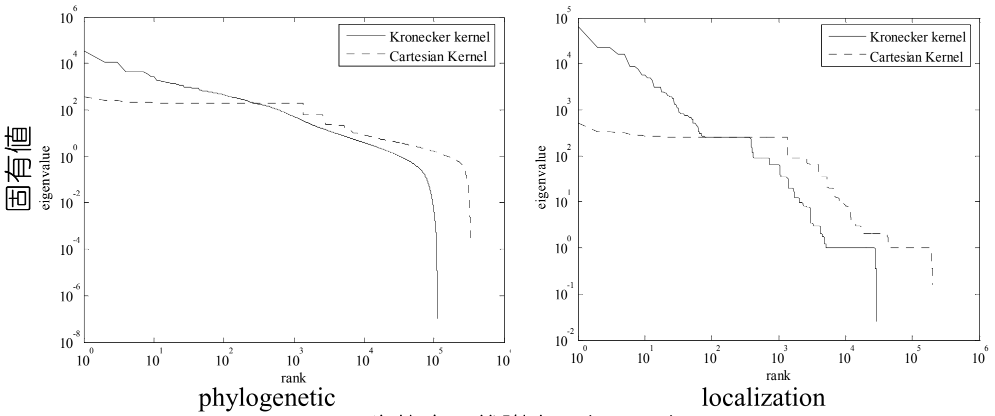
[定理] インスタンス間のカーネル行列 $\mathbf{K}^{(1)}$ と $\mathbf{K}^{(2)}$ の固有値の集合がそれぞれ $\{\lambda_i^{(1)}\}$と$\{\lambda_j^{(2)}\}$のとき、これらの

- クロネッカー積 $\mathbf{K}^{(2)} \otimes \mathbf{K}^{(1)}$ の固有値は$\{\lambda_i^{(1)} \lambda_j^{(2)}\}$ （固有値の積）
- クロネッカー和 $\mathbf{K}^{(2)} \oplus \mathbf{K}^{(1)}$ の固有値は$\{\lambda_i^{(1)} + \lambda_j^{(2)}\}$ （固有値の和）となる

クロネッカーカーネルとカルテシアンカーネルでは固有値の分布に違いが見られます

- 上位の固有値はクロネッカーカーネルの方が大きい
- 下位の固有値はカルテシアンカーネルの方が大きい

（つまり、クロネッカーカーネルの方が固有値の減衰が速い）

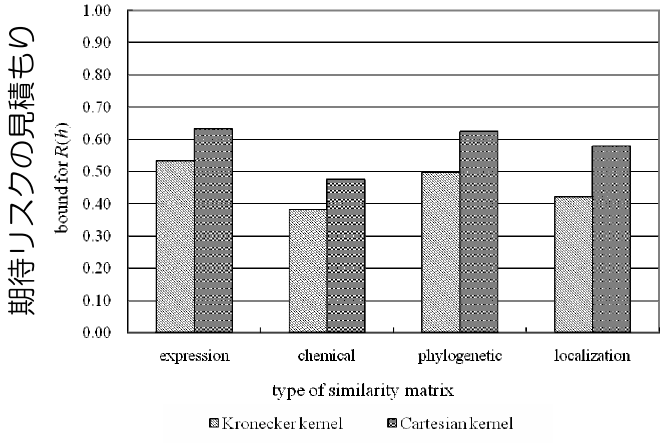


phylogenetic　localization

生体内の代謝ネットワーク

理論的な解析からは「クロネッカーカーネルの方がカルテシアンカーネルよりもよい」ことが示唆されます

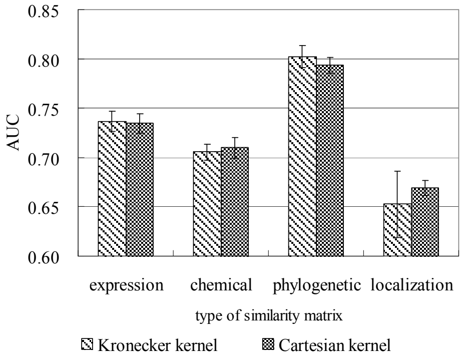
- 理論的な見積もりでは、クロネッカーカーネルの方が期待リスクが小さくなります
  - 但し、このバウンドの値自体はあまりタイトではありません



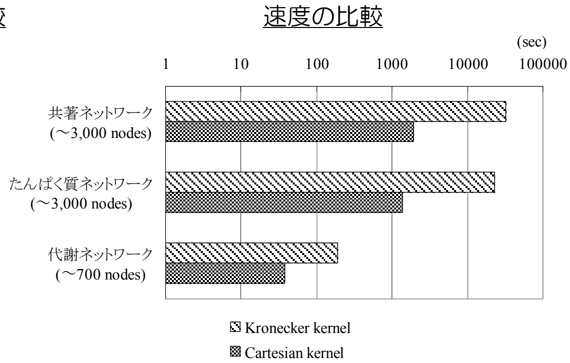
しかし、実験的には、カルテシアンカーネルは十分に高精度で、かつ、高速です

- ２つのペアワイズカーネルの予測性能差は、実験的にはあまりありません
- 一方で、カルテシアンカーネルは遥かに高速です

代謝ネットワークにおける精度の比較



速度の比較



※オンラインSVMを用いた

結論：新しいペアワイズカーネル「カルテシアン・カーネル」は、既存のペアワイズカーネルの有望な代替法です

- あたらしいペアワイズカーネル「カルテシアン・カーネル」を提案しました
  - 既存のペアワイズカーネルと比較して、計算が簡単
  - カーネル行列もスパース
- 理論的には、既存のペアワイズカーネルよりも悪いが、実験的には、あまり変わらないことを確認しました