変数間因果関係に関する
リレーショナルデータマイニングへの取り組み

鷲尾　隆，猪口明博，清水昌平，
河原吉伸，Steffen Rendle

大阪大学　産業科学研究所 / JST ERATO
知能推論研究分野
URL http://www.ar.sanken.osaka-u.ac.jp/

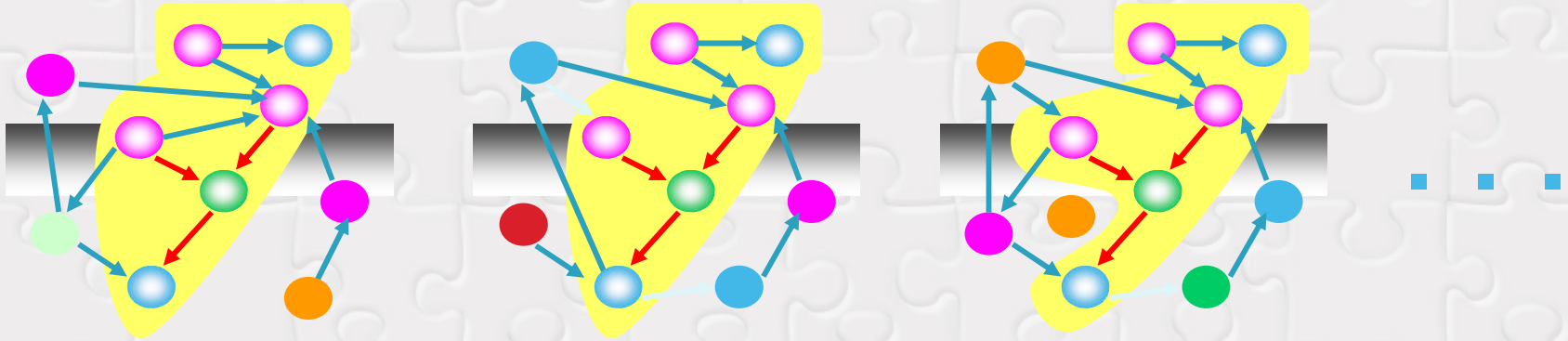
概要

1. 構造からのデータマイニング
   グラフマイニングを用いた変数間依存性マイニング
   (猪口明博，井元清哉，樋口知之等)
2. データからの構造マイニング
   非ガウス性に基づく変数間依存性マイニング
   （清水昌平，河原吉伸，Aapo Hyvarinen等）
3. 大規模テーブルからの予測モデルマイニング
   テンソル分解によるランキング予測
   （Steffen Rendle, Lars Schmidt-Thieme等）
4. １つの研究展望
   次世代リレーショナルデータマイニングに向けて

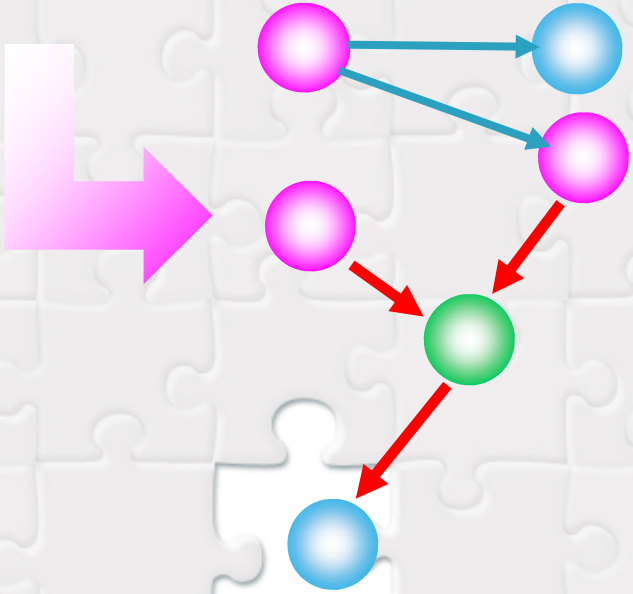
概要

1. 構造からのデータマイニング
   グラフマイニングを用いた変数間依存性マイニング
   （猪口明博，井元清哉，樋口知之等）
2. データからの構造マイニング
   非ガウス性に基づく変数間依存性マイニング
   （清水昌平，河原吉伸，Aapo Hyvarinen等）
3. 大規模テーブルからの予測モデルマイニング
   テンソル分解によるランキング予測
   （Steffen Rendle, Lars Schmidt-Thieme等）
4. １つの研究展望
   次世代リレーショナルデータマイニングに向けて

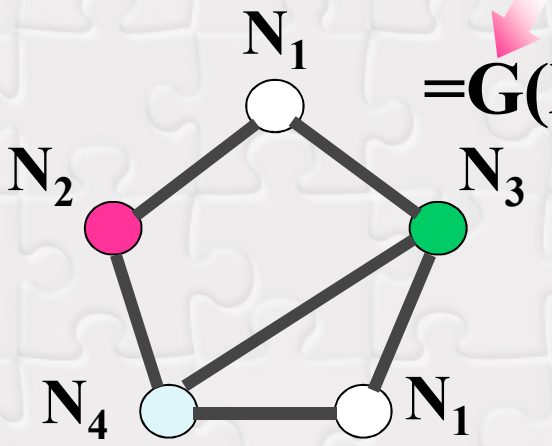
グラフマイニングの基礎
(Inokuchi et al., PKDD00, FI05)



- 対象データはグラフ事例の集合
- データ中の頻出部分グラフを探索



グラフマイニングの基礎



グラフ

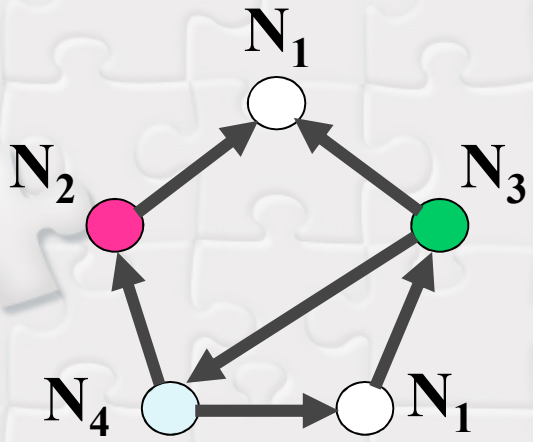
=G(X)

無向グラフ

	N_1	N_1	N_2	N_3	N_4
N_1	0	0	1	1	0
N_1	0	0	0	1	1
N_2	1	0	0	0	1
N_3	1	1	0	0	1
N_4	0	1	1	1	0

=X

隣接行列



有向グラフ

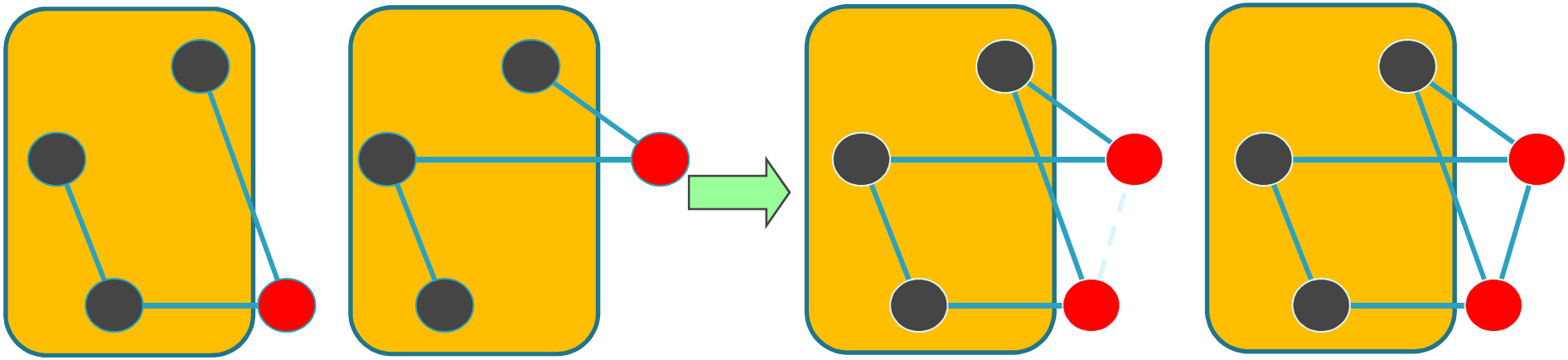
	N_1	N_1	N_2	N_3	N_4
N_1	0	0	0	0	0
N_1	0	0	0	1	0
N_2	1	0	0	0	0
N_3	1	0	0	0	1
N_4	0	1	1	0	0

=X

グラフマイニングの基礎

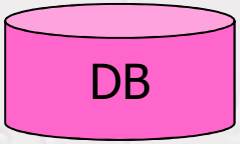
- 2つの多頻度グラフX_k, Y_kを結合することによって多頻度グラフ候補Z_{k+1}を生成する.

$$X_k=\begin{pmatrix} X_{k-1} & x_1 \\ x_2^T & 0 \end{pmatrix} \quad Y_k=\begin{pmatrix} X_{k-1} & y_1 \\ y_2^T & 0 \end{pmatrix} \Rightarrow Z_{k+1}=\begin{pmatrix} X_k & y_1 \\ & z_{k,k+1} \\ y_2^T \quad z_{k+1,k} & 0 \end{pmatrix}$$



グラフマイニングの基礎

マイニングアルゴリズム



最小支持度=2

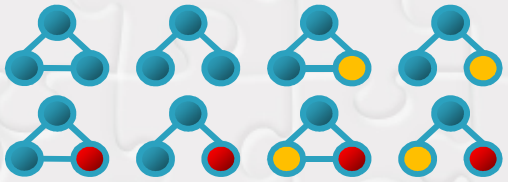
1-candidate subgraph

1-frequent subgraph

2-candidate subgraph

2-frequent subgraph

3-candidate subgraph



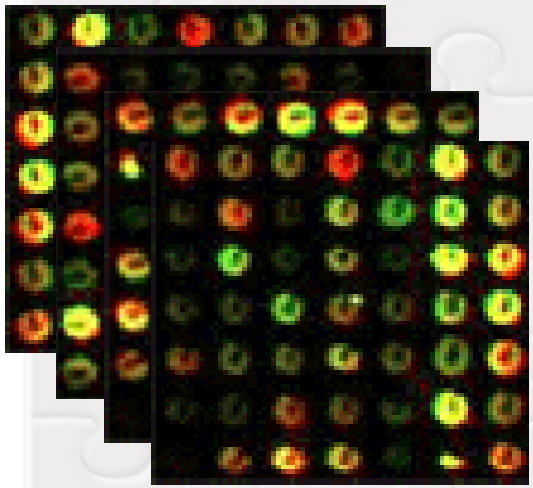
3-frequent subgraph

代表的
グラフマイニング
アルゴリズム
参照：http://hms.liacs.nl/

gSpan, Gaston
最速でよく使われる。

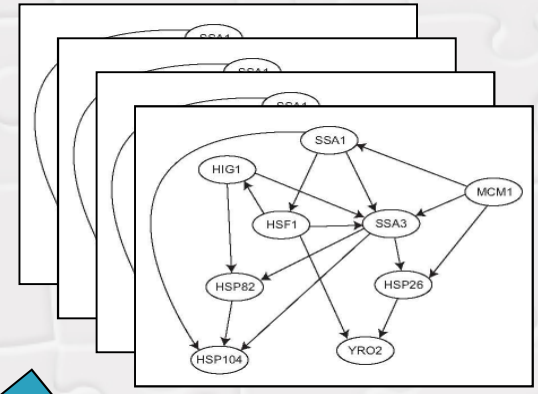
AGM, AcGM
初期のアルゴリズム。
(Inokuchi et al.,
PKDD2000, FI2005)

変数間依存性マイニングへの応用
遺伝子発現に関する統計的因果推論
（猪口，井元，鷲尾，樋口等 2007）



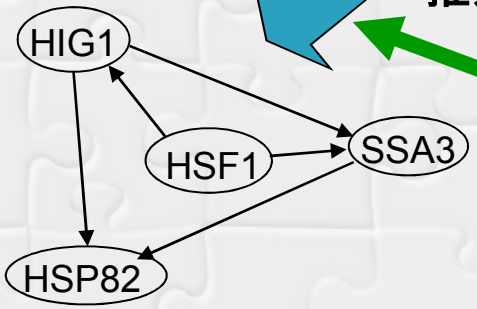
マイクロアレー
遺伝子発現
プロフィールデータ

ベイジアンネットワーク
ノンパラメトリック回帰
モデルを用いた
因果構造モデリング



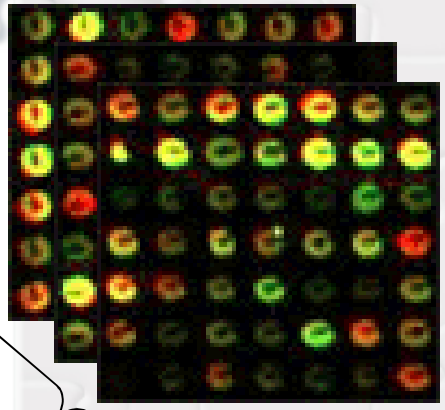
推定因果構造ネットワーク

グラフマイニングを用いた
頻出因果構造ネットワーク導出



不変的因果構造の発掘

統計的因果推論モデルの構築



n

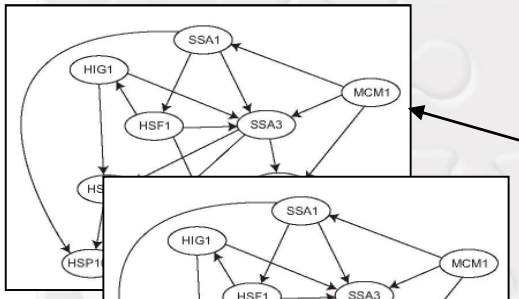
n枚のマイクロアレー $\{x_1, \cdots, x_n\}$

Bayesian network and nonparametric hetero-scedastic regression model (Imoto et al. 2003)

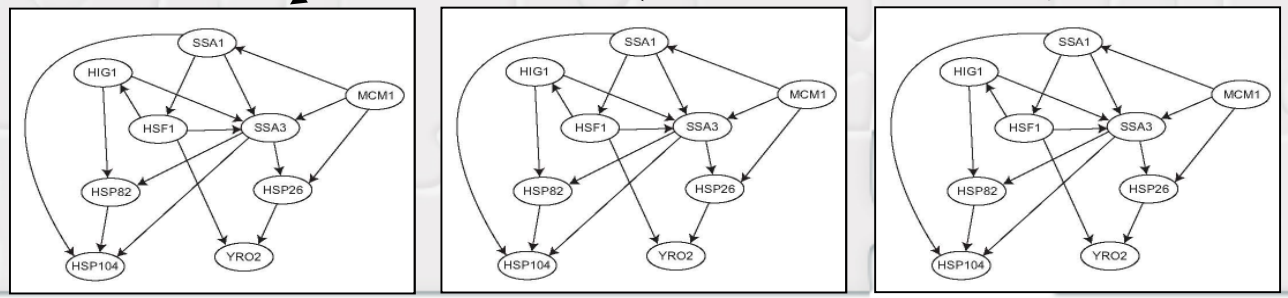
$$x_{ij} = m_{j1}\left(p_{i1}^{(j)}\right) + \cdots + m_{jq_j}\left(p_{iq_j}^{(j)}\right) + e_{ij}$$

$$f\left(x_i; \Theta_G\right) = \prod_{i=1}^{p} f_j\left(x_{ij} \mid p_{ij}; \Theta_j\right)$$

最大事後確率(MAP)解の最良優先探索



5000の
極大事後確率局所解



Process Go Termとグラフマイニングの適用

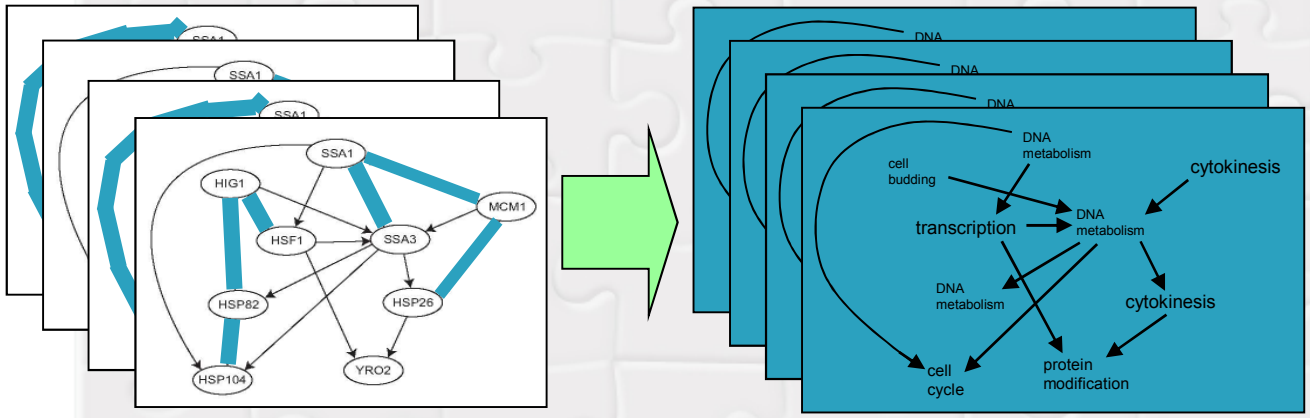
Gene Ontology (GO) Term：遺伝子の機能を表現する記述子
1つの遺伝子に
Process, Function, Component
の3つの側面からGO Termと呼ばれる記述を割り当て

例）遺伝子名称：YMR043W

Component	:	nucleus
Function	:	DNA binding
Process	:	DNA metabolism

→ 33種類の Process GO Term

データ中の各遺伝子固有名をProcess GO Termに付け替える.

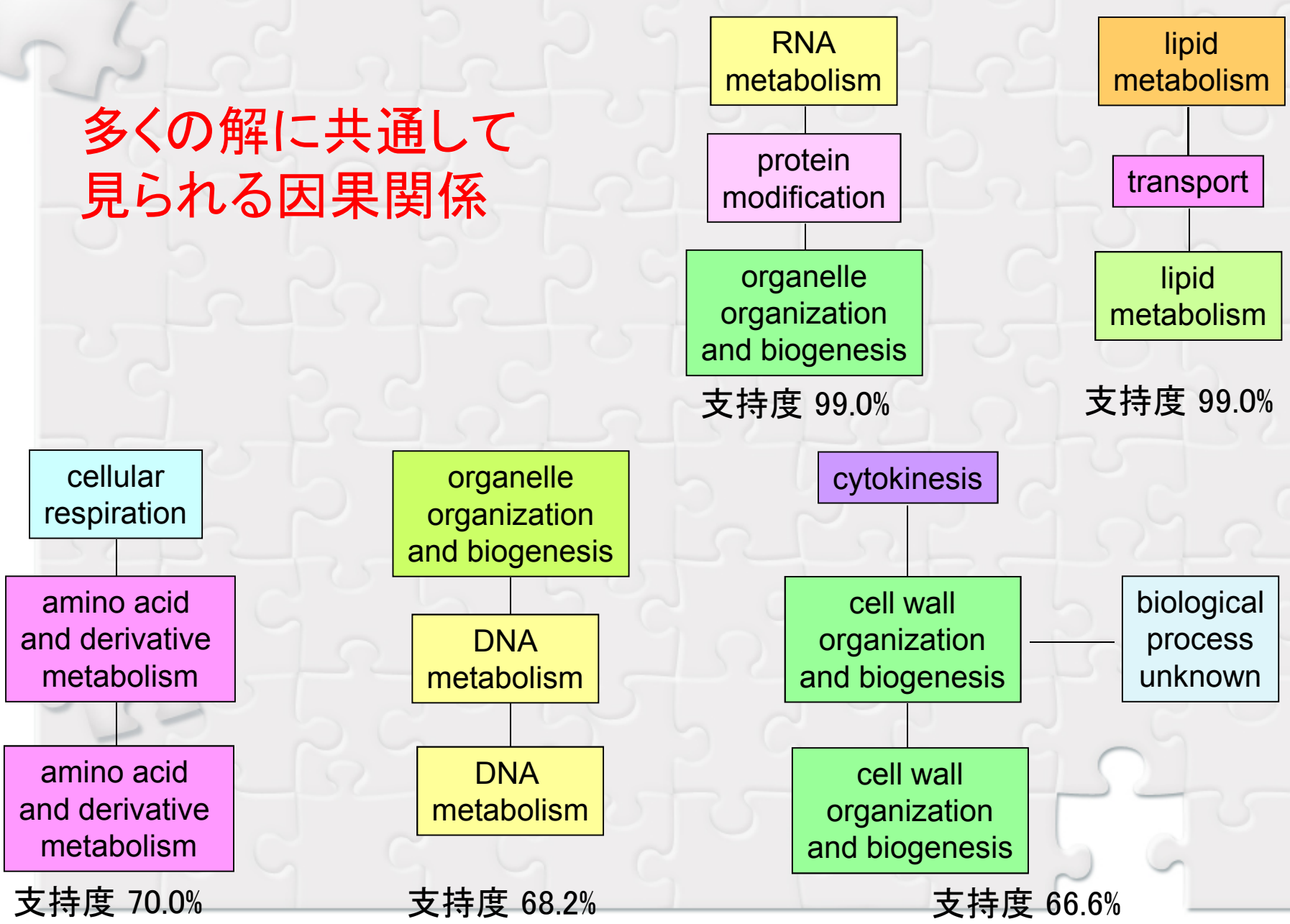


同一ラベル頂点が複数存在し，
頂点や辺をラベルで識別不能.
トポロジーを含むグラフとして扱う.

多頻度連結誘導部分グラフマイニングAcGM手法を適用

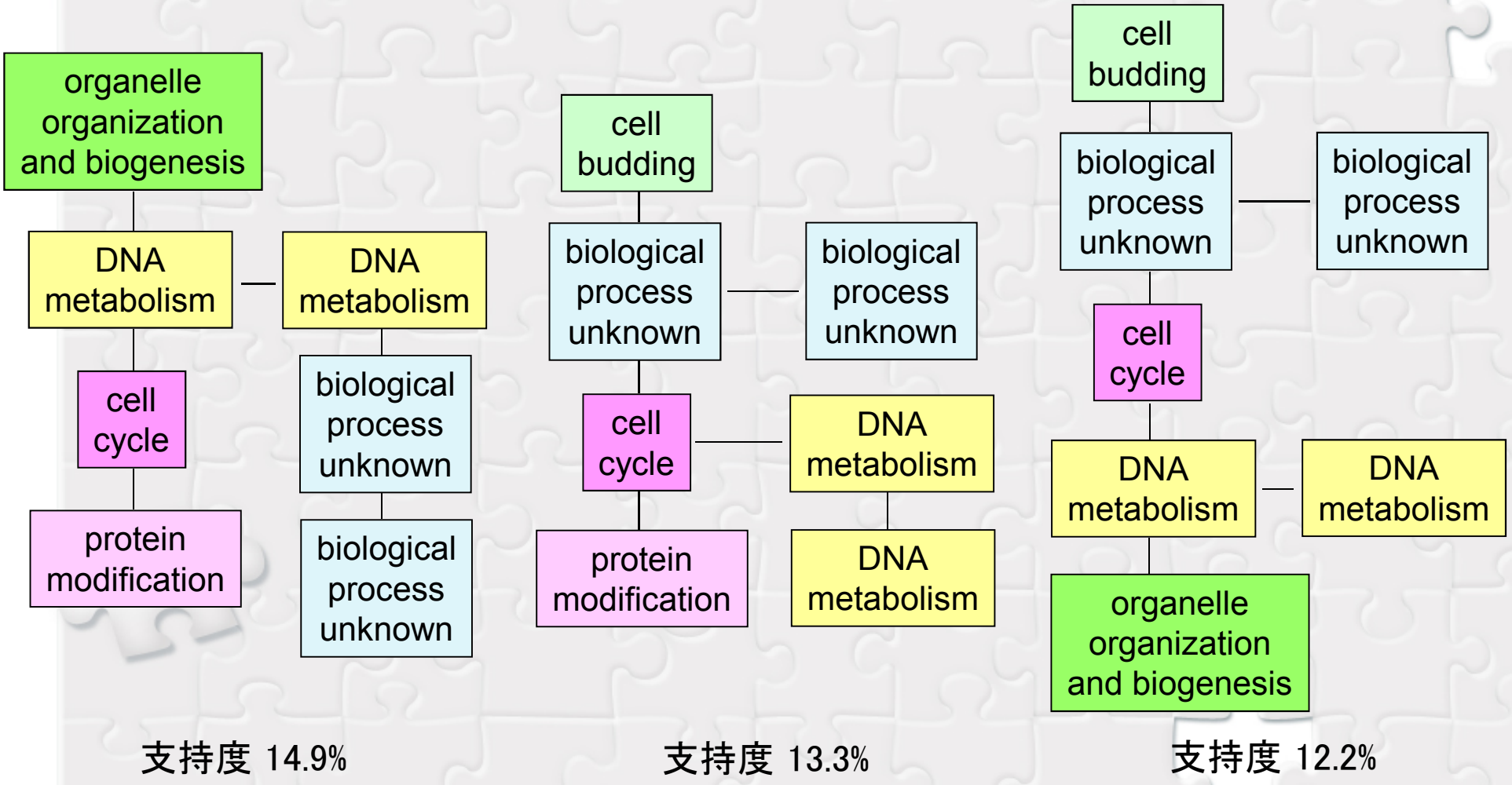
グラフマイニング結果（1）

多くの解に共通して
見られる因果関係



グラフマイニング結果（2）

一部の解に見られる比較的多くの遺伝子間の因果関係



グラフマイニングを用いた変数間依存性マイニング
まとめ

- 多変数データにベイジアンネットワークを適用し，多数の変数間依存関係ネットワークの候補を導出．
- 多数の候補集合にグラフマイニングを適用し，実際に依存関係が成立している可能性の高い変数間依存関係の部分ネットワークを導出．

問題点

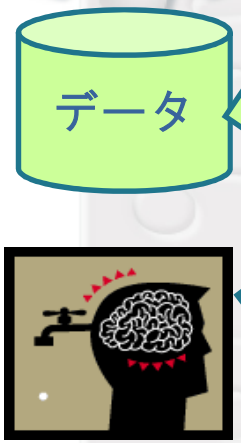
ベイジアンネットワークが大量の候補解を残し，かつそれらの候補が多様過ぎて，共通した明確な変数間依存関係の部分ネットワークが得られない．

概要

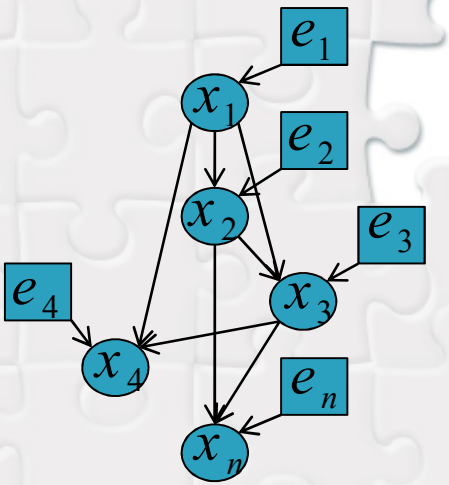
1. 構造からのデータマイニング
   グラフマイニングを用いた変数間依存性マイニング
   （猪口明博，井元清哉，樋口知之等）
2. データからの構造マイニング
   非ガウス性に基づく変数間依存性マイニング
   （清水昌平，河原吉伸，Aapo Hyvarinen等）
3. 大規模テーブルからの予測モデルマイニング
   テンソル分解によるランキング予測
   （Steffen Rendle, Lars Schmidt-Thieme等）
4. １つの研究展望
   次世代リレーショナルデータマイニングに向けて

統計的因果推論（グラフィカルモデリング）の概観

因果構造の学習（構造学習）や因果強度の学習（パラメータ学習）



データと人間の背景知識から変数間の決定関係を推定



因果関係を表す有向非巡回グラフ
DAG (Directed Acyclic Graph)

共分散構造分析

$$
\begin{aligned}
x_1 &= e_1 \\
x_2 &= b_{21}x_1 + \cdots + e_2 \\
x_3 &= b_{31}x_1 + b_{32}x_2 + \cdots + e_3 \\
x_4 &= b_{41}x_1 + b_{43}x_3 + \cdots + e_4 \\
&\vdots \\
x_n &= b_{n2}x_2 + b_{n3}x_3 + \cdots + e_n
\end{aligned}
$$

互いに独立な非観測外乱

変数間決定関係を線形モデルで同定

ベイジアンネットワーク

$$
\begin{aligned}
&p(x_1) \\
&p(x_2 \mid x_1) \\
&p(x_3 \mid x_1, x_2) \\
&p(x_4 \mid x_1, x_3) \\
&\vdots \\
&p(x_n \mid x_1, x_3, \cdots, x_{n-1})
\end{aligned}
$$

変数間決定関係を確率モデルで同定

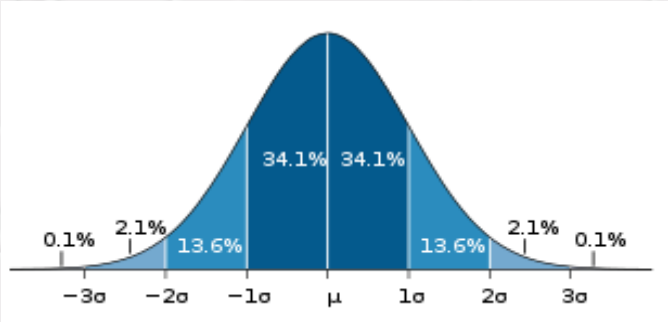
バイオインフォマティクスやマーケティングなどで多用される解析手法

従来の統計的因果推論の問題点

- 従来の統計的因果推論は相関や偏相関など
  ２次の統計量を用いて変数間の因果的順序を導出。
  非観測外乱項や確率密度がガウス分布と仮定。

$$C(x_1, x_2) = \frac{\sum_{i=1}^{n} x_1 x_2}{\sqrt{\sum_{i=1}^{n} x_1^{\,2}}\sqrt{\sum_{i=1}^{n} x_2^{\,2}}} = 0$$

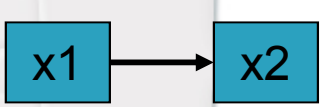
⇒ x_1とx_2は無関係



- しかし、このガウス性の仮定（あるいは近似）により、一般に識別不可能な因果構造が存在する.
  (Pearl, 2000; Spirtes et al. 2000):

Model 1　$x_1 := b_{12} x_2 + e_1$

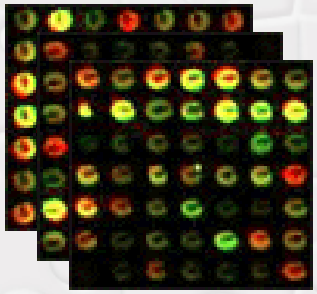
Model 2　$x_2 := b_{21} x_1 + e_2$

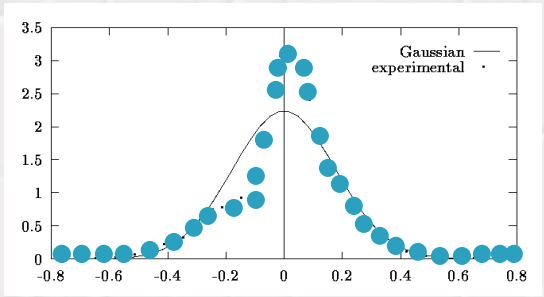


非観測外乱項や確率密度の非ガウス性に着目

(Shimizu et al., UAI05, JMLR06, UAI09)

- "Many observed data are considerably non Gaussian." (Micceri, 1989; Hyvarinen et al. 2001)





- 非ガウス成分を考えることで、変数間の因果構造は識別可能となる。

$$
\begin{aligned}
x_1 &= e_1 \\
x_2 &= b_{21}x_1 + \cdots + e_2 \\
x_3 &= b_{31}x_1 + b_{32}x_2 + \cdots + e_3 \\
x_4 &= b_{41}x_1 + b_{43}x_3 + \cdots + e_4 \\
&\vdots \\
x_n &= b_{n2}x_2 + b_{n3}x_3 + \cdots + e_n
\end{aligned}
$$

Model 1 $x_1 := b_{12}\, x_2 + e_1$ x1 ← x2

Model 2 $x_2 := b_{21} x_1 + e_2$ x1 → x2

互いに独立な非ガウスの非観測外乱

LiNGAMモデル (Linear Non-Gaussian Acyclic Model)

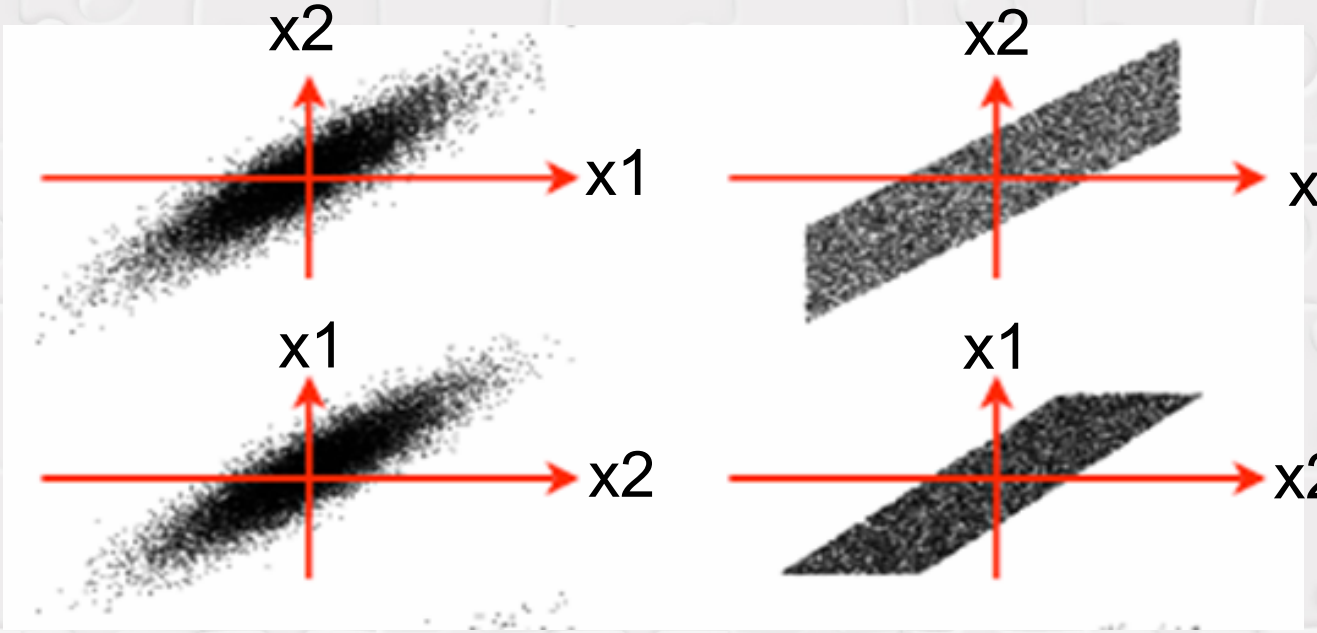
ガウス性 v.s. 非ガウス性

Model 1

$$x_2 := \beta\, x_1 + e_2$$

Model 2

$$x_1 := \beta\, x_2 + e_1$$



Gaussian

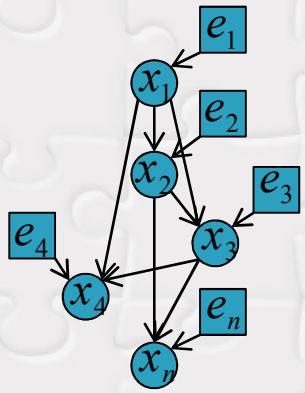
Non-Gaussian

提案手法：線形非正規非巡回モデル

Linear Non-Gaussian Acyclic Model (LiNGAM model)

(Shimizu et al. UAI05)

- 構造方程式モデル (Structural Equation Model:SEM)



DAG (Directed Acyclic Graph) Structure

$$
\begin{aligned}
x_1 &= e_1 \\
x_2 &= b_{21}x_1 + \cdots + e_2 \\
x_3 &= b_{31}x_1 + b_{32}x_2 + \cdots + e_3 \\
x_4 &= b_{41}x_1 + b_{43}x_3 + \cdots + e_4 \\
&\vdots \\
x_n &= b_{n2}x_2 + b_{n3}x_3 + \cdots + e_n
\end{aligned}
$$

- 各観測変数 x_i は連続確率変数.
- 変数間の因果関係は線形で有向非巡回グラフ(DAG) で表される.
- 各外乱 e_i は互いに独立で非ガウス揺らぎを持つ.

$$x_i = \sum_{k(j)<k(i)} b_{ij}x_j + e_i \quad \text{or} \quad \mathbf{x} = \mathbf{Bx} + \mathbf{e}$$

- 行列Bは行と列の同時入れ替えによって下三角化される.

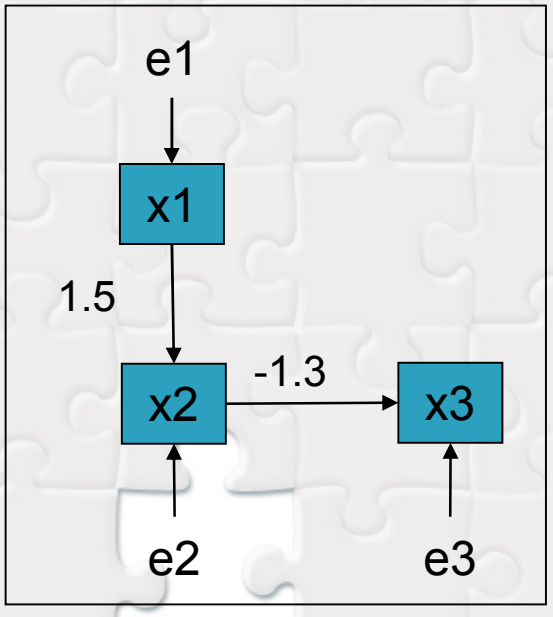
具体例

- ３変数モデル

$$x_1 = e_1$$
$$x_2 = 1.5x_1 + e_2$$
$$x_3 = -1.3x_2 + e_3$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 1.5 & 0 & 0 \\ 0 & -1.3 & 0 \end{bmatrix}}_{\mathbf{B}} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}$$

- 変数の因果的順序
  - $k(1) = 1, k(2) = 2, k(3) = 3$
  - x2 はx1から影響を受けるが，x3からは影響を受けない．
- 外乱
  - 外乱x1は外乱e1に等しい．x1を外生変数という．外生変数は最も上流の変数である．
  - e2とe3は付加外乱あるいは誤差 である．



提案手法(Shimizu et al.,UAI09)

変数間因果の判定(２変数の例)

i) $x_1 (= e_1)$ が上流である.

$x_2 := b_{21} x_1 + e_2$

x_1 and e_2 are independen t.

x_2 and e_2 are NOT independen t.

x_1とx_2の回帰式

$$r_2 = x_2 - \frac{\mathrm{cov}(x_2, x_1)}{\mathrm{var}(x_1)} x_1 = e_2$$

x_1 and r_2 are independent.

ii) x_1 が上流でない.

$x_1 := b_{12} x_2 + e_1$

x_2 and e_1 are independen t.

x_1 and e_1 are NOT independen t.

x_1とx_2の回帰式

$$r_2 = x_2 - \frac{\mathrm{cov}(x_2, x_1)}{\mathrm{var}(x_1)} x_1$$

$$= \left\{ 1 - \frac{b_{12}\,\mathrm{cov}(x_2, x_1)}{\mathrm{var}(x_1)} \right\} x_2 - \frac{\mathrm{cov}(x_2, x_1)}{\mathrm{var}(x_1)} e_1$$

x_1 and r_2 are NOT independent.

提案手法
独立性を測る尺度

- 厳密にはあらゆる非線形関数f,gの非線形相関ゼロを確認せねばならないが、現実には無理。

独立性判定：非線形相関係数

$$C_{fg}(x_1, r_2) = \frac{\sum_{i=1}^{n} f(x_1) g(r_2)}{\sqrt{\sum_{i=1}^{n} f(x_1)^2} \sqrt{\sum_{i=1}^{n} g(r_2)^2}} = 0$$

（あらゆる非線形関数f,gについて）

- 以下の非線形相関により近似的に変数とその残差の独立性を検定評価

$$C_{fg}(x_1, r_2) \cong \left|\mathrm{corr}\left\{x_1, g(r_2)\right\}\right| + \left|\mathrm{corr}\left\{g(x_1), r_2\right\}\right| \quad (g = \tanh)$$

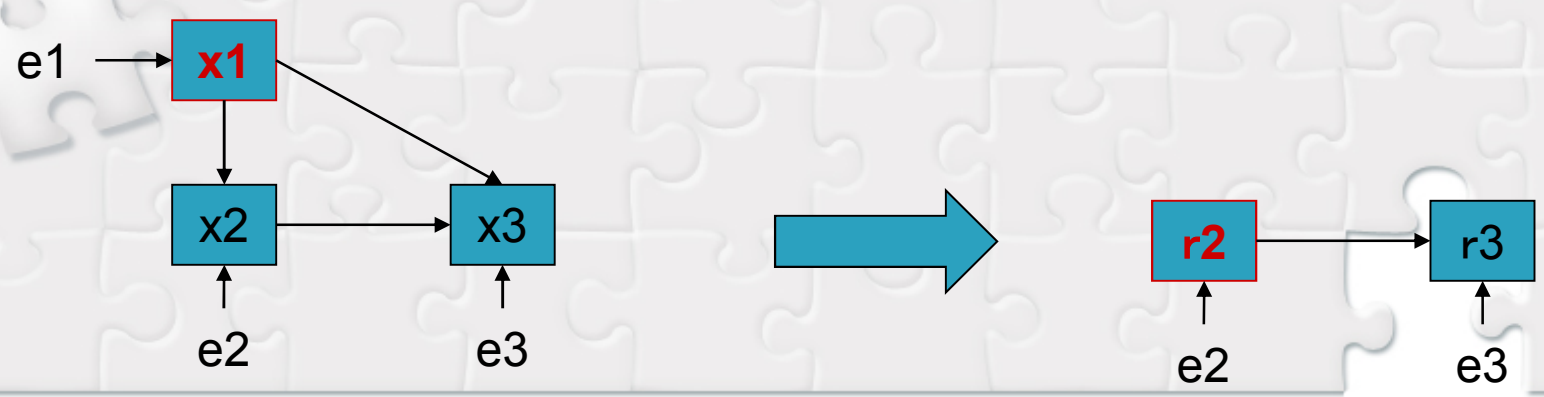
モデルの前提がデータで成立していれば，非線形相関の近似誤差を除いて，この方法は必ず正しい唯一の因果構造を同定することが証明されている．

DirectLiNGAM: 直接的因果推論法

- 基本原理:

1. 前掲の回帰式法より最上流の外生変数を探す.
   - 例えば $x_1 \rightarrow \{x_2, x_3\}$ が分かる.
   - 回帰により最上流のx1の成分を他の変数データから除く.
2. “外生”の回帰残差を探す.
   - 例えば $r_2 \rightarrow r_3$, *i.e.*, $x_2 \rightarrow x_3$.

   残差の因果順序と元の変数の因果順序は一致する.
3. 最終的に $x_1 \rightarrow x_2 \rightarrow x_3$ が導かれる.

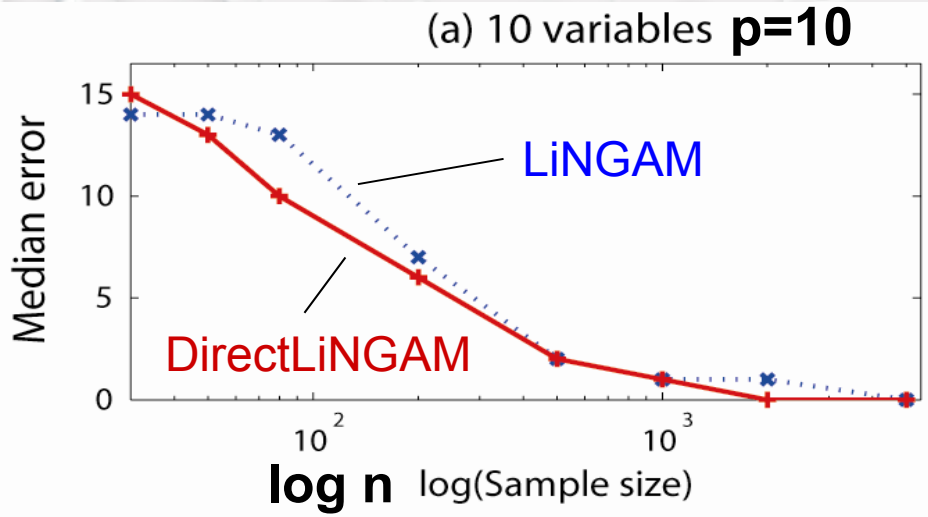


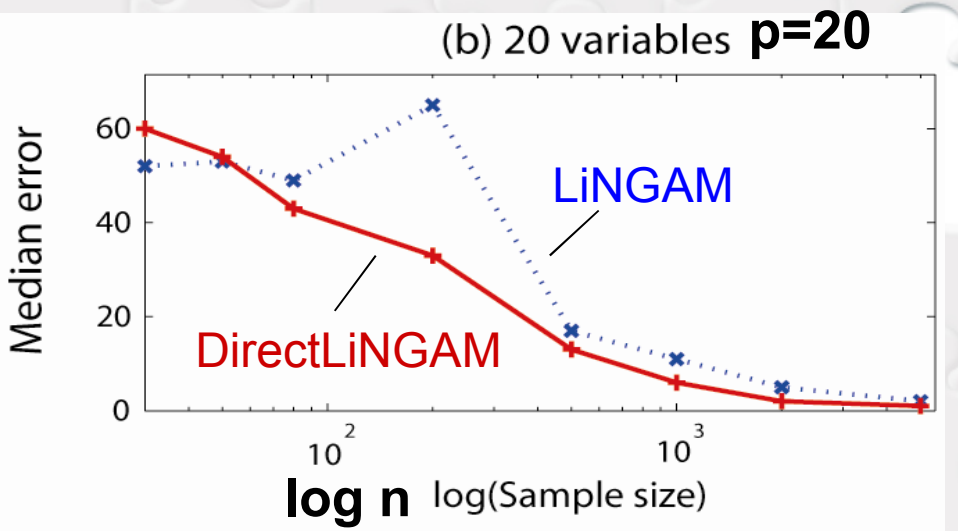
DirectLiNGAM:まとめ

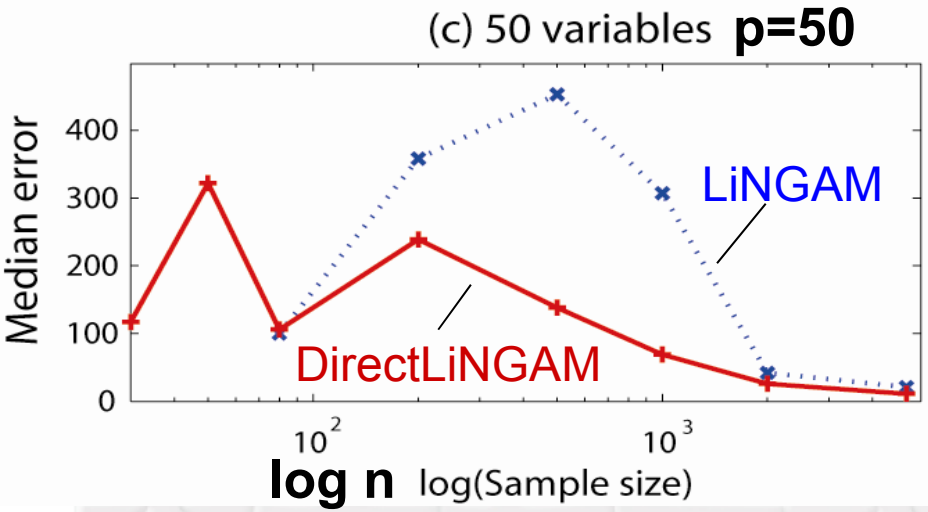
- 知っていること（あるいは仮定）
  - データXは$\mathbf{x} = \mathbf{B}\mathbf{x} + \mathbf{e}$によって生成されている.
- 知らないこと
  - 変数間の結合係数: b_{ij}
  - 因果的順序: k(i),
  - 各外乱:e_i.
- データXのみ観測されている.

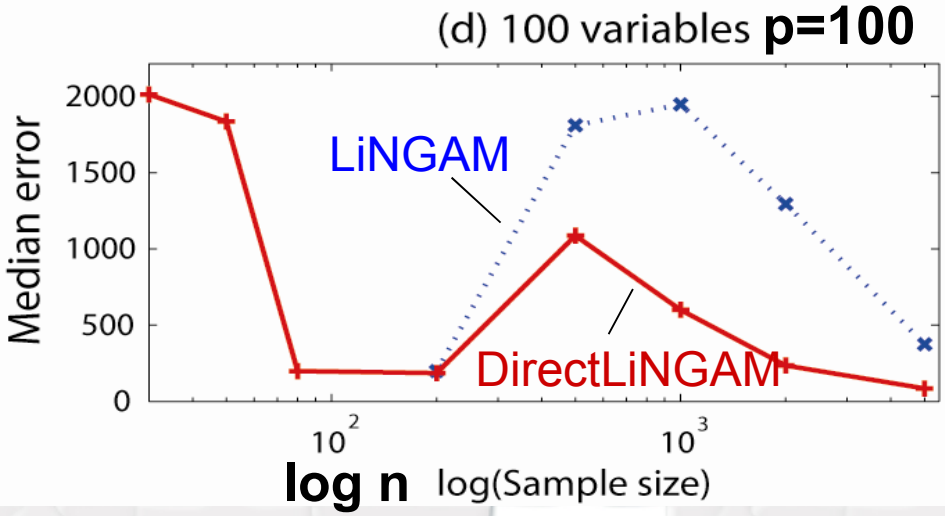
- データXのみからBとk,eが導びかれる！

標本サイズv.s.非ゼロ要素数（誤差）







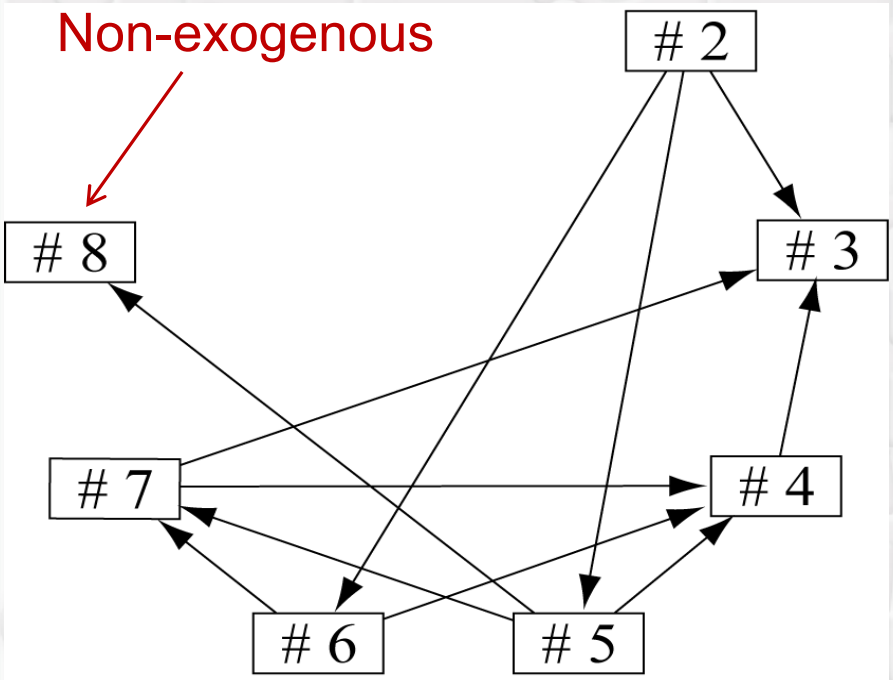


n = 30, 50, 80, 200, 500, 1000, 2000

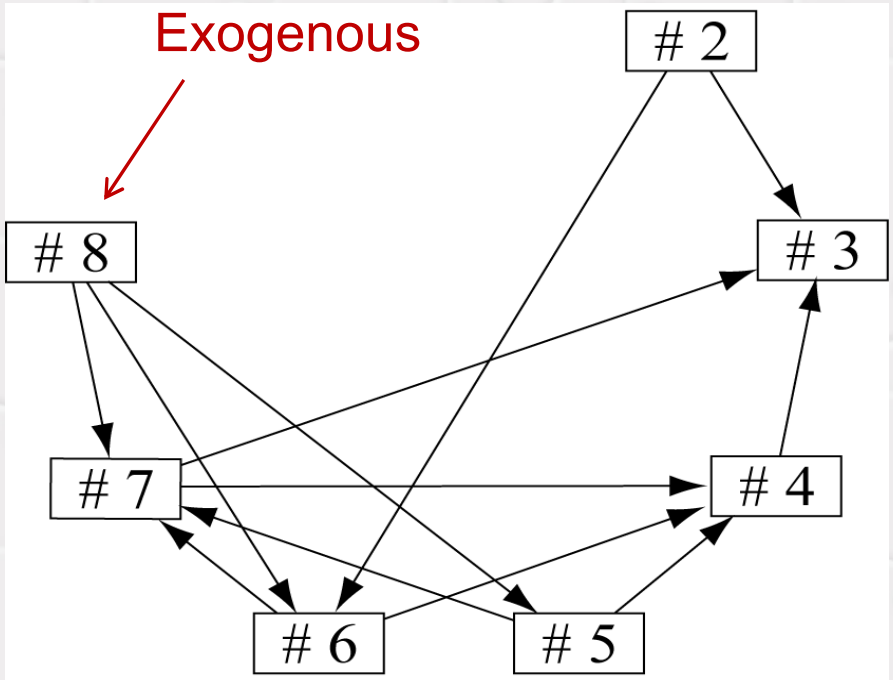
変数が100個程度のデータでも正確に因果構造を同定可能

脳波Magneto-encephalography data から導いた脳波源の影響関係

- 各手法により推定された因果構造:

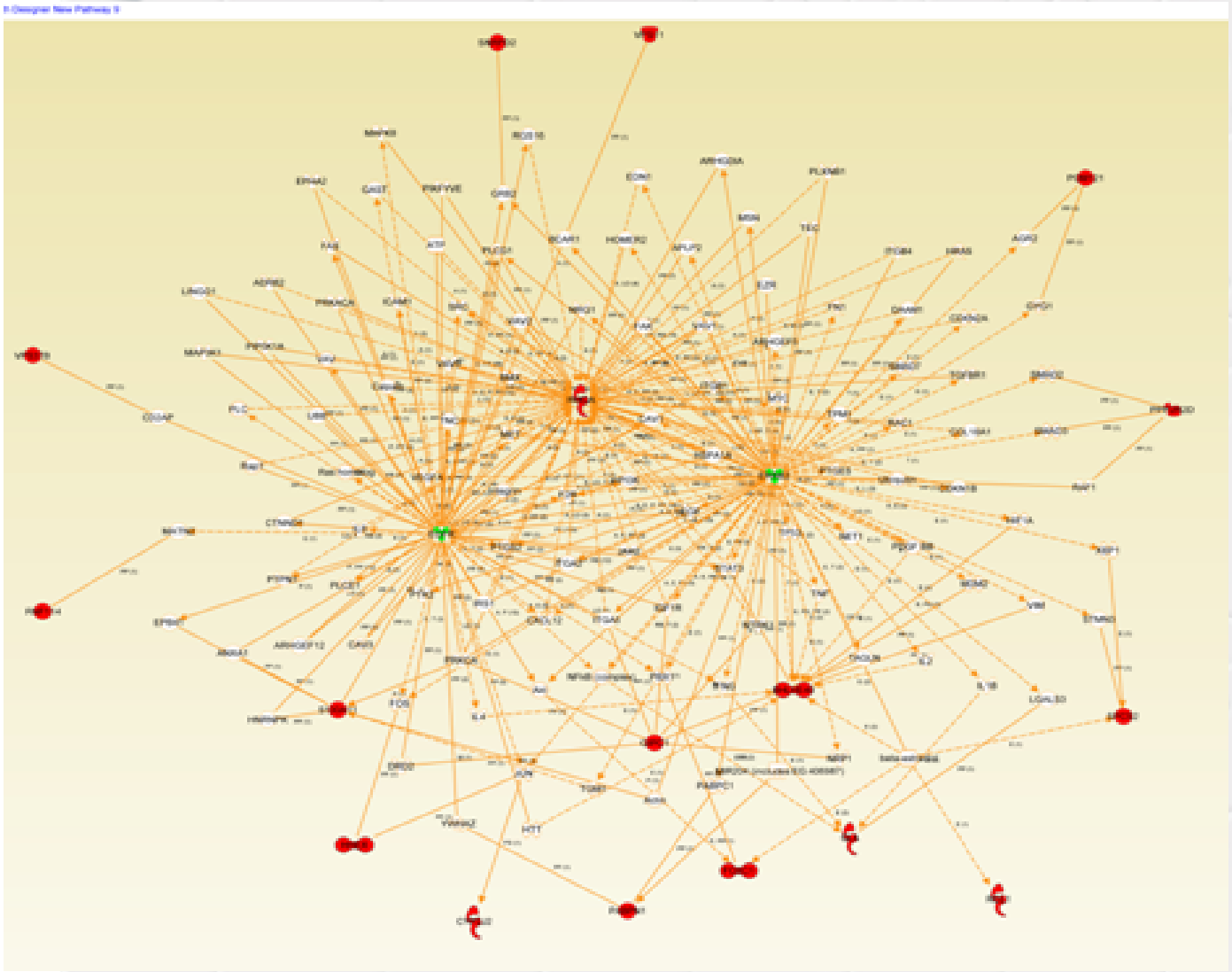


(Original) LiNGAM



DirectLiNGAM (proposed)

非ガウス性に基づく外生変数同定結果と遺伝子データベースから得た乳がん遺伝子発現因果ネットワーク



外生変数同定手法
eggFinder
(Shimizu and Sogawa et al. JSAI10)

+

Ingenuity Pathways Database

非ガウス性に基づく変数間依存性マイニング まとめ

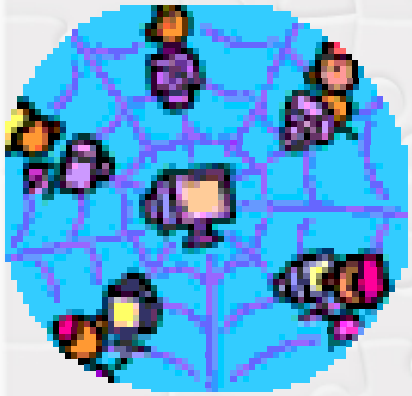
- 独立な外乱の非ガウス性を利用するLiNGAMモデルの導入により，変数間依存性の非識別性問題を克服.
- ２変数間の回帰分析と非線形相関係数に基づく独立性判定に基づき，変数間の依存性に関する一意な順序付けを行うDirect LiNGAMの提案
- 多変数間の因果解析・依存性マイニングの新たな方法論が確立しつつある.

概要

1. 構造からのデータマイニング
   グラフマイニングを用いた変数間依存性マイニング
   （猪口明博，井元清哉，樋口知之等）
2. データからの構造マイニング
   非ガウス性に基づく変数間依存性マイニング
   （清水昌平，河原吉伸，Aapo Hyvarinen等）
3. 大規模テーブルからの予測モデルマイニング
   テンソル分解によるランキング予測
   （Steffen Rendle, Lars Schmidt-Thieme等）
4. １つの研究展望
   次世代リレーショナルデータマイニングに向けて

テンソル分解によるランキング予測
(Rendle S. et al. SIG-KDD09, ECML-PKDD09, UAI09)
研究背景

- サービス，ネットワーク，センサ群など，大規模データを生成するシステムの増加



例）ネット販売サイトにおいて，顧客集団が過去に商品群を検索するのに用いたキーワードタグから，個別の顧客に商品毎の適切なキーワードタグを付けて提示したい。

カバン
バッグ
ビジネス
……

じょうろ
水やり
園芸
……

Tag Recommendation

顧客，商品毎に適切な順にタグをランキングして，上位タグを提示する.

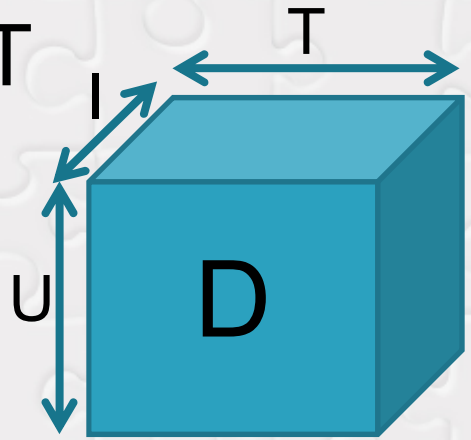
テンソル分解によるランキング予測

- 問題設定

  U:顧客集合，I:商品集合，T:タグ集合

- 購買記録のデータキューブ:U×I×T

- この購買記録から各要素の条件毎にタグの適切さの程度を推定したい。



例）|U|=300万人，|I|=100万点，|T|=2万語⇒一般に非常にスパース

顧客Aの商品4のタグがfならA，4，f-要素を1，

該当データがない欠測要素を0としてとすると．．．

＊殆どの要素はゼロ⇒極端なImbalance Dataとなる．

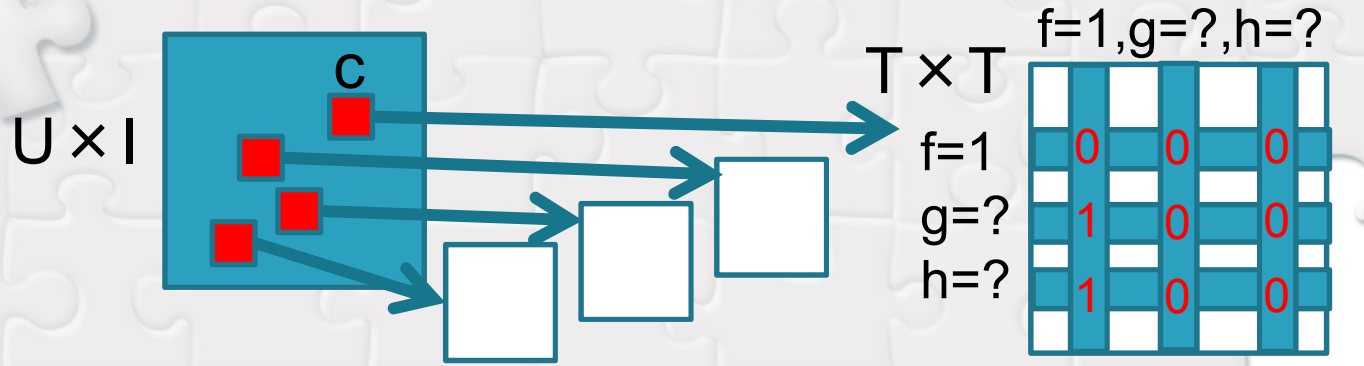
＊そもそも欠測要素は，記録がないだけで適切さがゼロなわけではない．

テンソル分解によるランキング予測

購買記録の実装

1. データキューブU×I×Tを
   コンテクストU×Iとランキング対象Tに区別する.
2. 各コンテクストc∈ U×I毎に 欠測要素の0を避けるために,
   ペア(ti,tj)∈T×Tで前者が後者より適切ti>tjなら1,
   ti<tjあるいはti ? tjなら0とする.

   例)コンテクストc=(顧客A, 商品4)∈ U×Iについて,
   タグfの要素は1 > タグgの要素は記録なし⇒g,h要素=1
   タグgの要素は記録なし ? タグhの要素は記録なし⇒g,h要素=0



テンソル分解によるランキング予測

Recommendation指標の推定

タグ同士の相対的適切さを保存する指標を探す.

1. 各コンテクストc∈I×Uにおいて各タグti∈T
   のRecommendation指標をy(c, ti|Θ)とする.
2. 各cにおいてti>tj⇔y(c, ti|Θ)>y(c, tj|Θ)を
   実現する指標y:I×U×T→Rのパラメータ Θ を探す.

パラメータ Θ の推定方法：MAP推定

全コンテクストにおけるタグ同士の相対的適切さの下で最尤パラメータを推定する.

$$\Theta = \operatorname{argmax}\ p(\Theta \mid I \times U; T \times T)$$
$$= \operatorname{argmax}\ p(I \times U; T \times T \mid \Theta)p(\Theta)$$ （ベイズの定理より）

テンソル分解によるランキング予測

Bayesian Context-aware Ranking Optimization (BCR-Opt)

$$\Theta = \text{argmax}\ p(I \times U; T \times T \mid \Theta) p(\Theta)$$

- Bernoulli分布とロジスティック分布の仮定

  $p(ti > tj \text{ under } c \mid \Theta) = 1 - p(tj > ti \text{ under } c \mid \Theta)$

  $p(ti > tj \text{ under } c \mid \Theta) = \sigma(y(c, ti) - y(c, tj))$ （σ：ロジスティック）

$$p(I \times U; T \times T \mid \Theta) = \prod_{(c,ti,tj) \in I \times U \times T \times T} p(ti > tj \text{ under } c \mid \Theta)^{\delta(ti>tj)}\, p(tj > ti \text{ under } c \mid \Theta)^{\delta(tj>ti)}$$

$$= \prod_{(c,ti,tj) \in I \times U \times T \times T} \sigma(y(c,ti) - y(c,tj))^{2\delta(ti>tj)}$$

- Prior $p(\Theta)$の正規分布の仮定

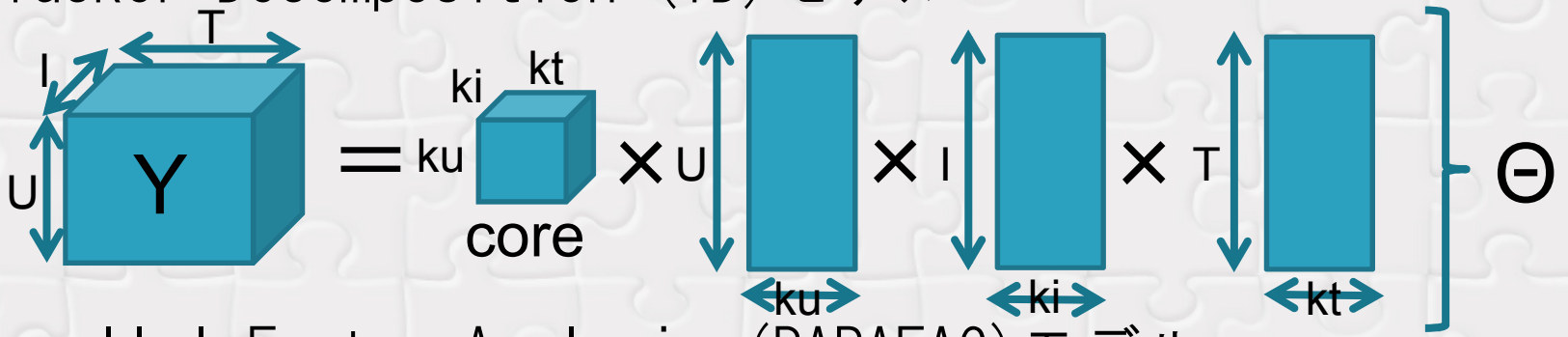
  $p(\Theta) = N(0, 1/\lambda)$ （λ：ハイパーパラメータ）

BCR-Learn：以上の基準の下で最尤パラメータΘを最大勾配法で推定する.

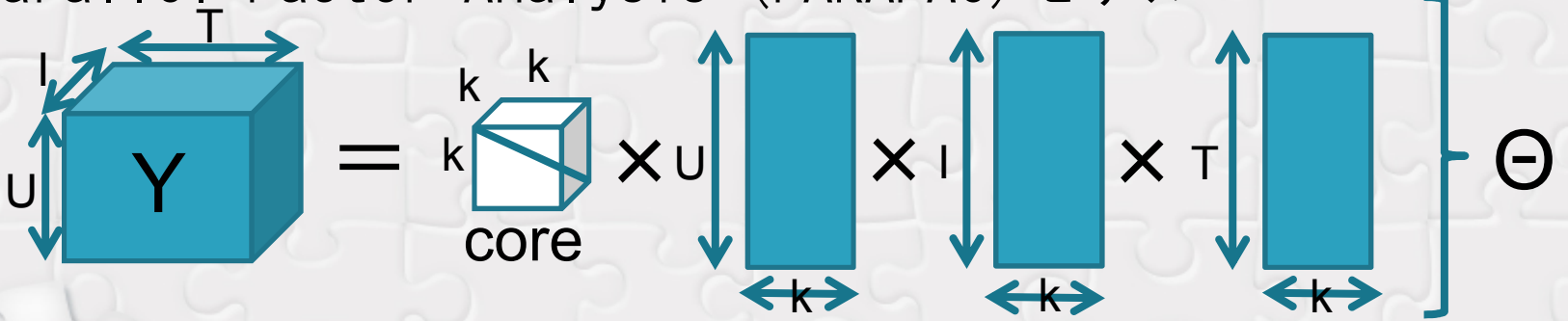
Recommendation指標 y(c, ti|Θ)のモデル

y:I×U×T→R　yを要素値とするテンソルの分解モデル

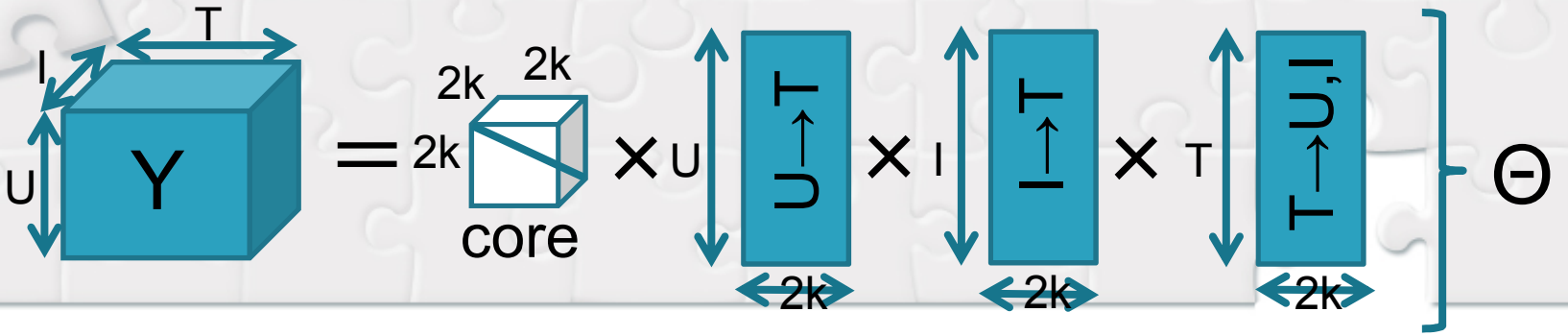
1. Tucker Decomposition (TD)モデル



2. Parallel Factor Analysis (PARAFAC)モデル



3. Pairwise Interaction Tensor Factorization (PITF)モデル



テンソル分解によるランキング予測

評価実験(Rendle et al. SIG-KDD09)

1. 使用データ

BibSonomy: ソーシャルブックマークシステムのデータ

Last.fm:イギリスのインターネットラジオ運営会社のSNSデータ

Data Set	Users |U|	Items |I|	Tags |T|	Records |S|
BibSonomy	116	361	412	10,148
Last.fm	2,917	1,853	2,045	219,702

2. 評価指標

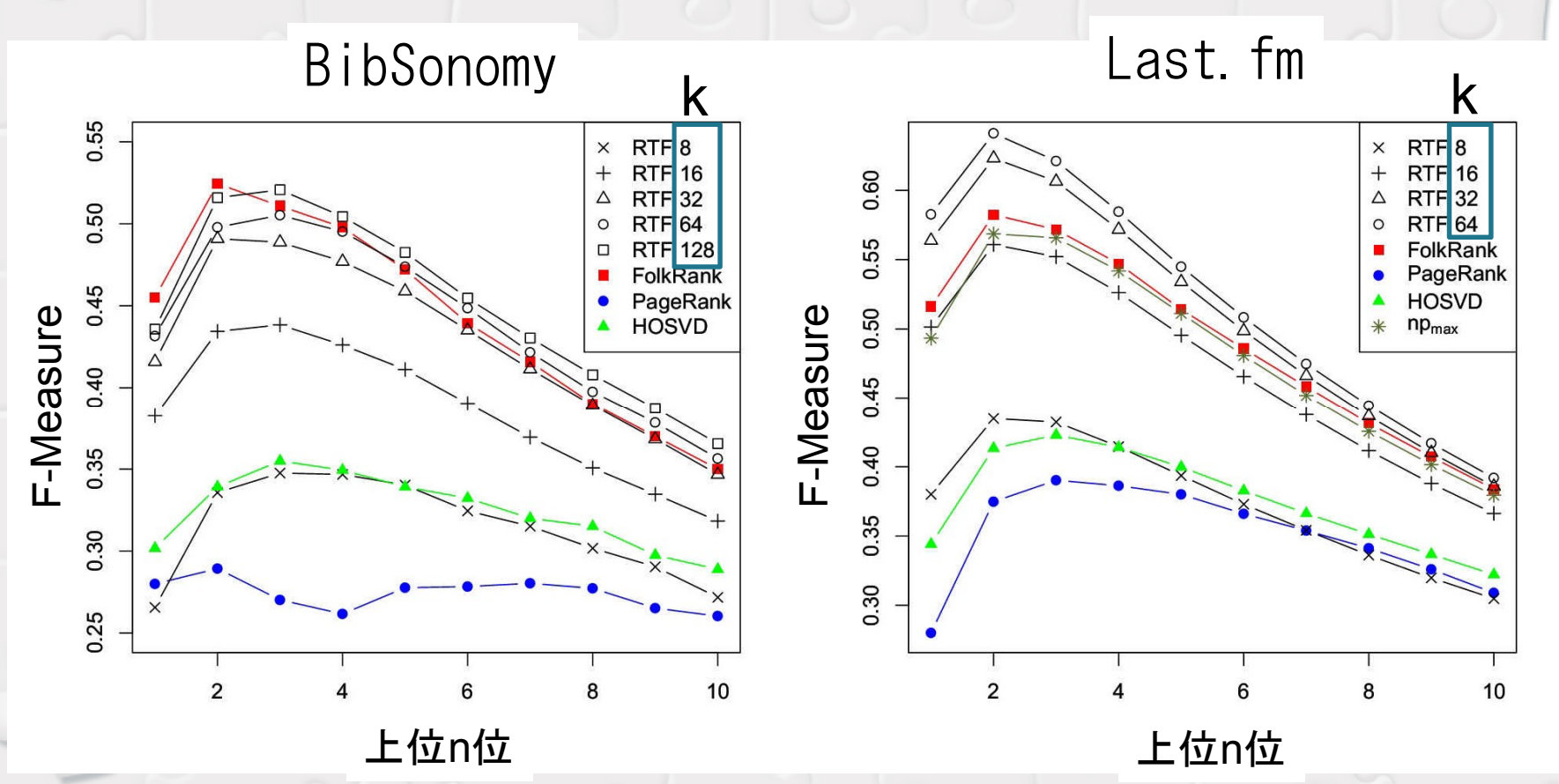
Sc:コンテキストcで真に使われたキーワードタグリスト,

Rc:コンテキストcで推定された上位n位キーワードタグリスト

$$\mathrm{Precision}(U\times I,n)=\frac{1}{|U\times I|}\sum_{c\in U\times I}\frac{|Sc\cap Rc|}{n},\mathrm{Recall}(U\times I,N)=\frac{1}{|U\times I|}\sum_{c\in U\times I}\frac{|Sc\cap Rc|}{|Sc|}$$

$$\mathrm{F-Measure}(U\times I,n)=\frac{2\mathrm{Precision}(U\times I,n)\mathrm{Recall}(U\times I,n)}{\mathrm{Precision}(U\times I,n)+\mathrm{Recall}(U\times I,n)}$$

テンソル分解によるランキング予測



BibSonomyとLast.fmに関する上位n位のRecommendationのF-Measure

テンソル分解によるランキング予測
まとめ

- 大規模でスパースなテーブル形式データから高精度な予測や推定を行うための一対比較法によるデータ実装方法を提案した。
- 更に上記データから目的変数に関するランキングを行う指標に関する確率モデルを提案した。
- この指標モデルとして適したテンソル分解モデルを提案した。
- この結果、現状のランキング手法よりも高精度な推定結果が得られた。

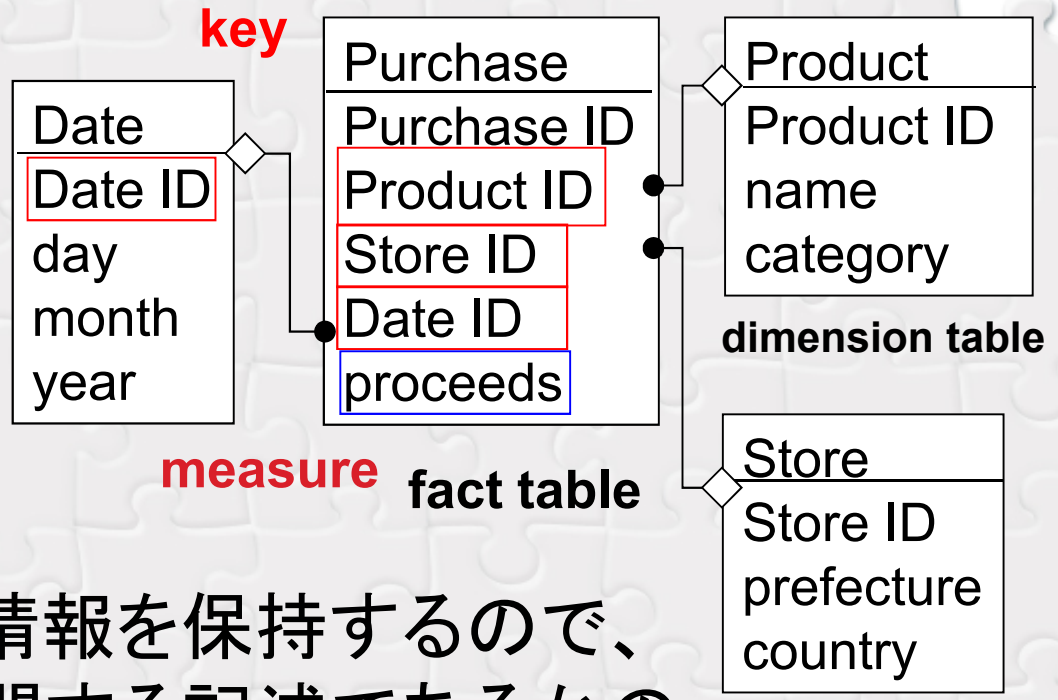
概要

1. 構造からのデータマイニング
   グラフマイニングを用いた変数間依存性マイニング
   （猪口明博，井元清哉，樋口知之等）
2. データからの構造マイニング
   非ガウス性に基づく変数間依存性マイニング
   （清水昌平，河原吉伸，Aapo Hyvarinen等）
3. 大規模テーブルからの予測モデルマイニング
   テンソル分解によるランキング予測
   （Steffen Rendle, Lars Schmidt-Thieme等）
4. １つの研究展望
   次世代リレーショナルデータマイニングに向けて

次世代リレーショナルデータマイニングへの展望

- リレーショナルデータベース

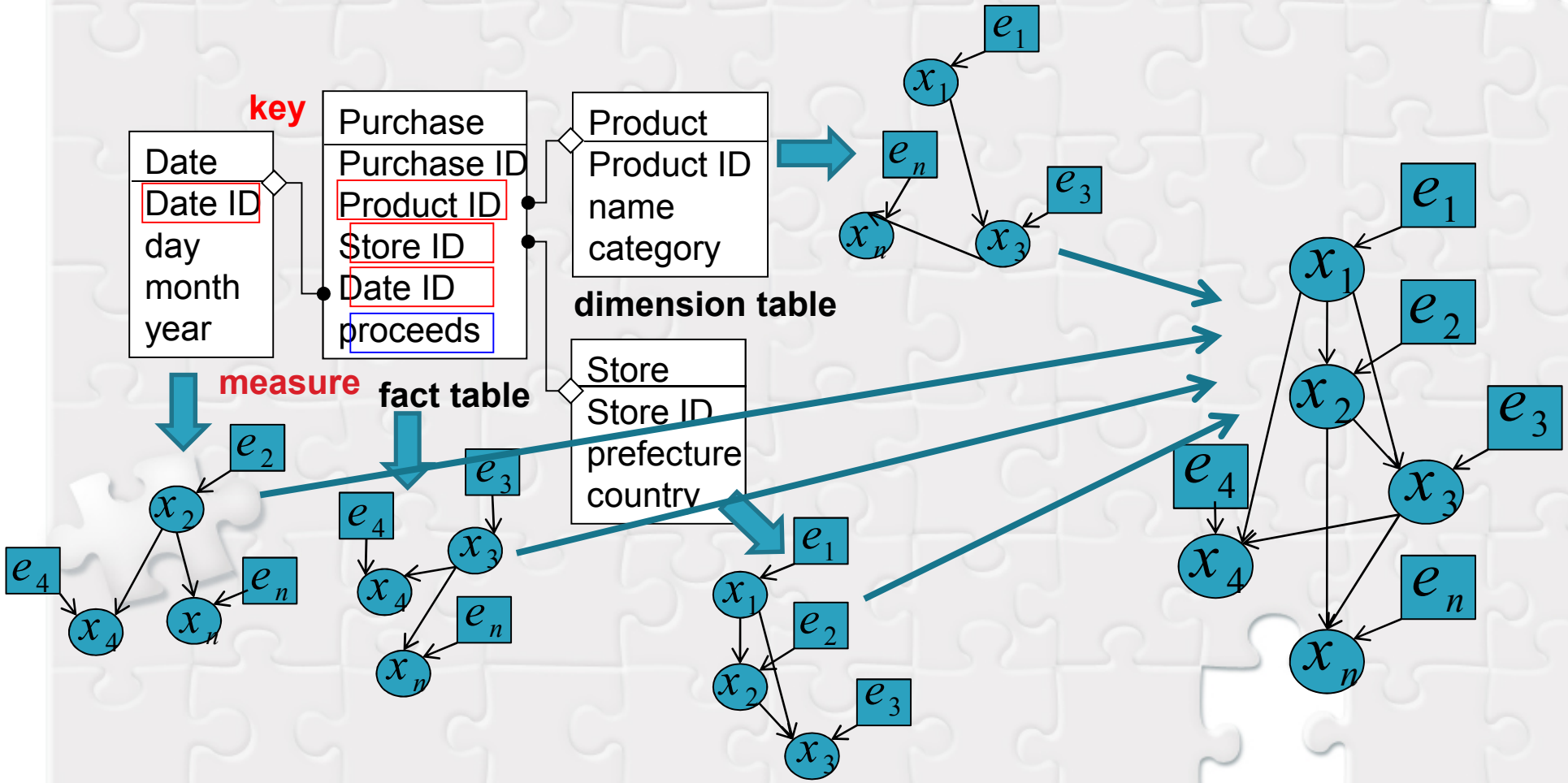
ID	Date	Product	Store
13	2008/2/8	drink	Tokyo
14	2008/2/9	snack	Osaka



表ごとにkeyに関する記述情報を保持するので、
結合(join)してしまうと何に関する記述であるかの
情報が失われてしまう。
⇒表の関係構造を保持したマイニングが望ましい。

次世代リレーショナルデータマイニングへの展望

- リレーショナルデータベース上での統計的因果推論



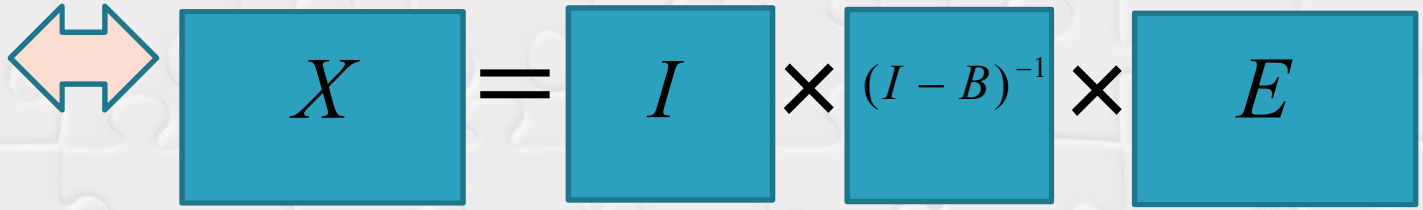
次世代リレーショナルデータマイニングへの展望

- リレーショナルデータベース上での
  テンソル分解による統計的因果推論？？

LiNGAMモデル

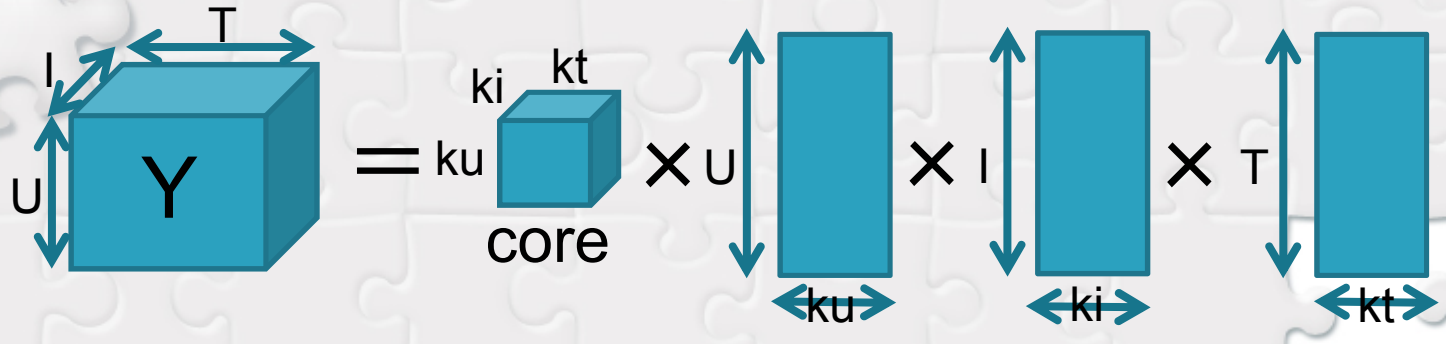
$$\mathbf{x}_i = \mathbf{B}\mathbf{x}_i + \mathbf{e}_i$$

$$(i = 1, \ldots, n)$$



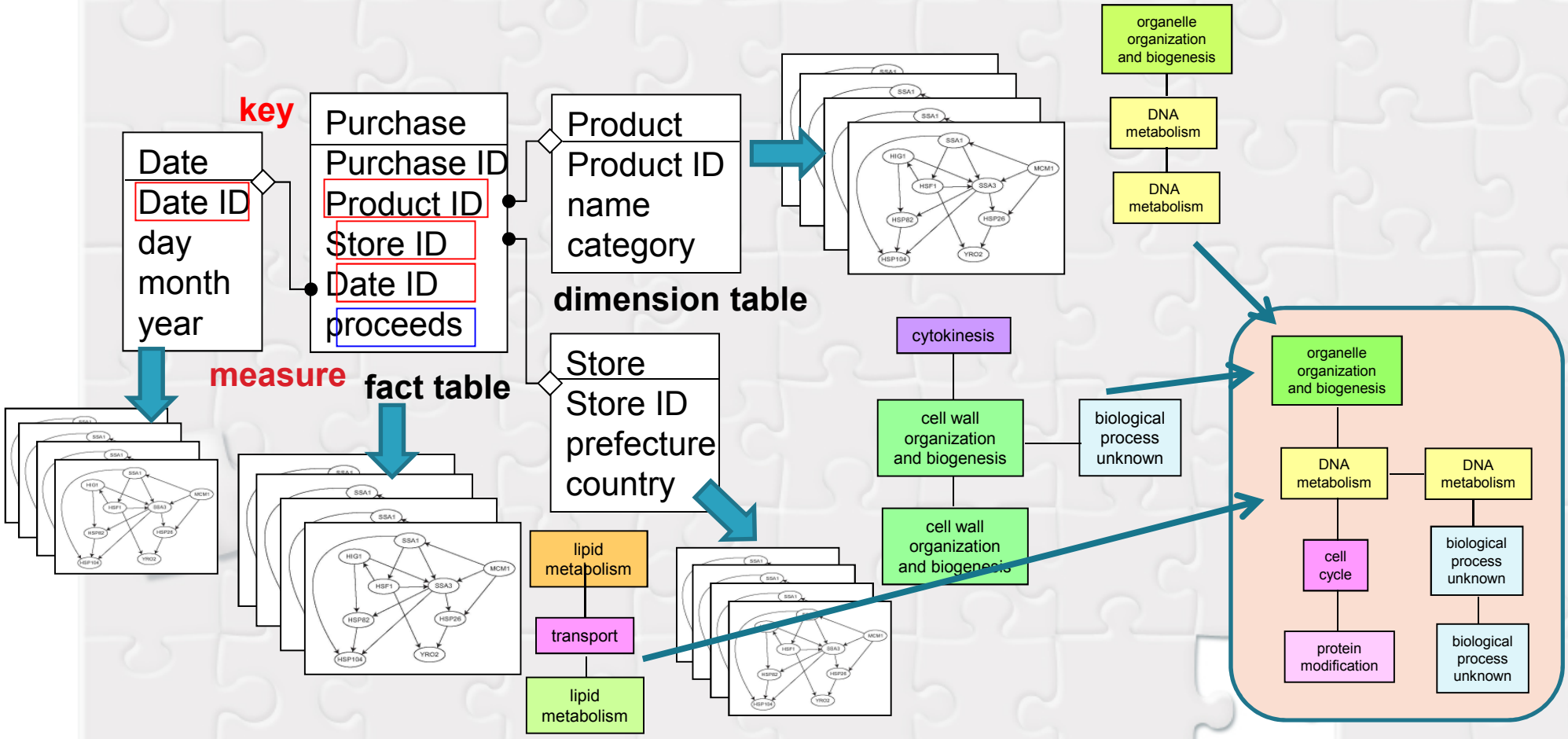
ICA

$$X = I \times (I - B)^{-1} \times E$$



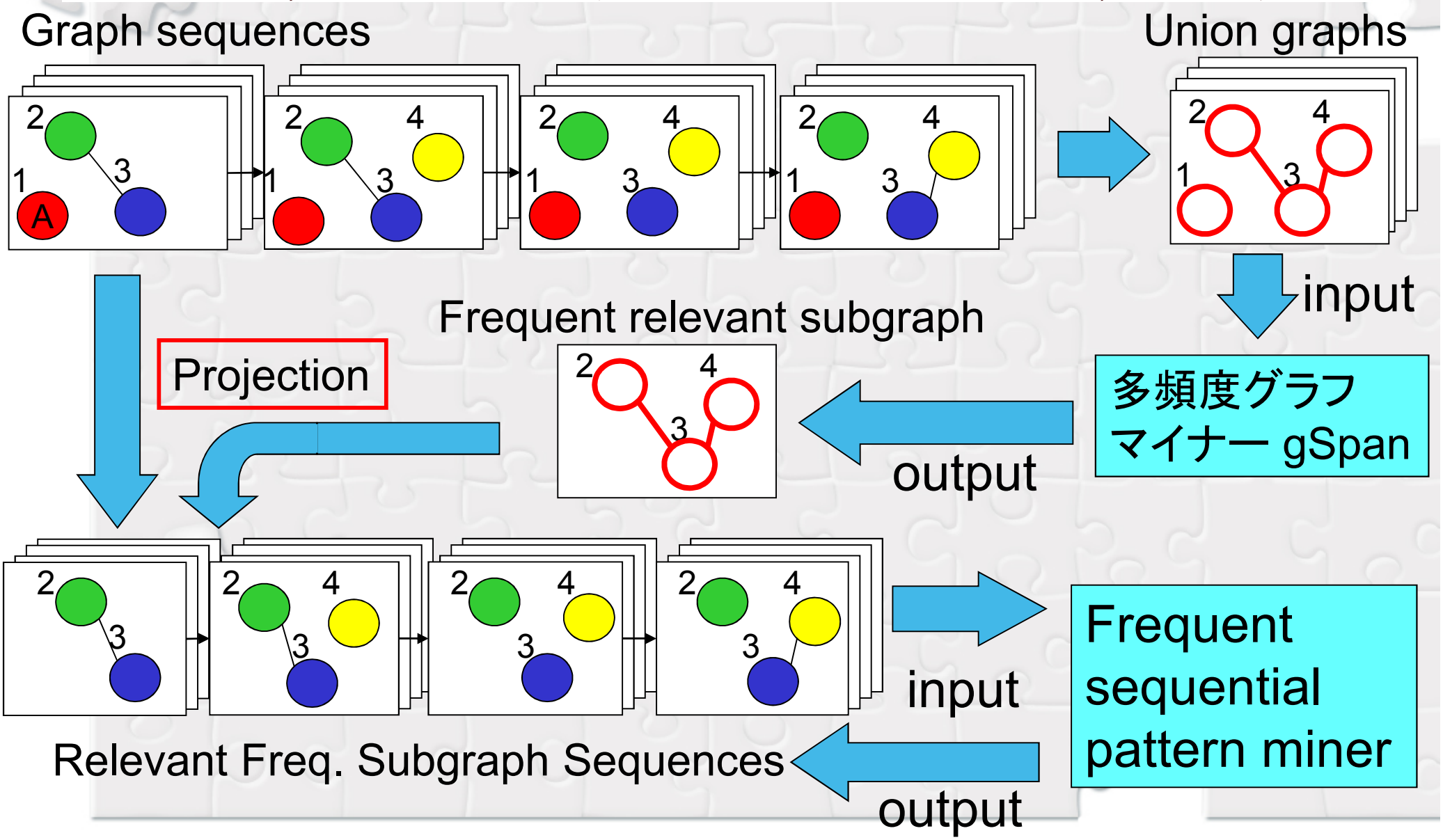
次世代リレーショナルデータマイニングへの展望

- リレーショナルデータベース上での
  リレーショナルなグラフ集合からのデータマイニング



多頻度グラフ系列のマイニング

GTRACE, FRISSMiner (Inokuchi at al. ICDM08, SDM10)



おわりに

- 構造からのデータマイニング
- データからの構造マイニング
- 大規模テーブルからのモデルマイニング

？　　リレーショナルデータマイニング　　？

現実の対象やそのデータは何らかの構造や関係と関っていることが多い．様々な展開が予想される．