決定グラフを用いた組合せ最適化問題に対する統一的解法の研究

石畠 正和（NTTコミュニケーション科学基礎研究）

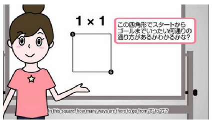
- 組合せ最適化とは、離散構造（グラフ、集合、論理などの離散的な対象）に関する最適化問題である。
- 組合せ最適化は様々な問題を包含するため、問題ごとに個別に解法が設計・実装されている。
- 統一的な解法が存在すれば、問題の変更や制約の追加に伴い解法を新たに設計・実装する必要がなくなる。
- 本研究プロジェクトでは決定グラフを用いて**組合せ最適化に対する統一的な解法**の構成を目指す。

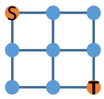
決定グラフ

離散構造を効率的に扱うための索引構造であり様々な基本演算を提供

離散構造は組合せ爆発を引き起こす

例）お姉さん問題（s-t パスの数え上げ）



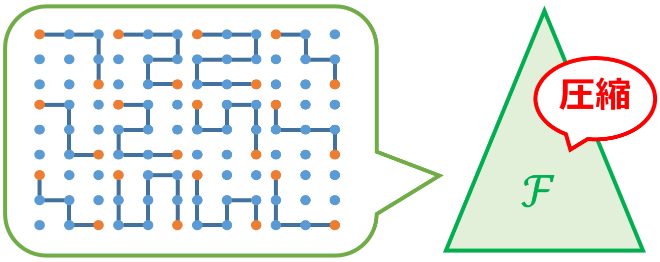


S から T に至るシンプルパスの数は？

- **グラフサイズに対して指数的に増加**
- **愚直に計算すると何億年もかかる**

索引構造（決定グラフ）

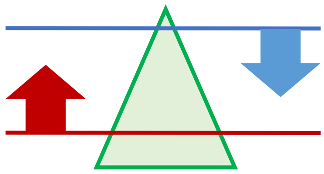
指数的に多い対象を圧縮表現



基本演算（動的計画法）

- 数え上げ
- 線形関数最大化
- 確率計算

などなど



プロジェクト成果

決定グラフを利用した高速・厳密な組合せ相関検定法（= **新たな演算**）

組合せ相関検定：「観測 x は**構造**に依存している？」という質問に答える

例）都道府県別の人口の増減は**地域性**を持つか？

帰無仮説：$x_1, \ldots, x_d$ は互いに独立

対立仮説：$\exists S \in \mathcal{F}, p(x_i \mid i \in S) \neq p(x_j \mid j \notin S)$

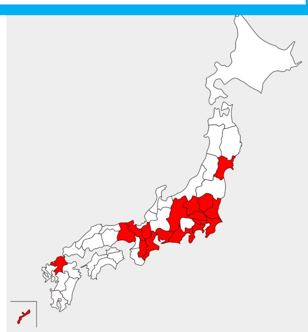
ある仮説パターン $S \in \mathcal{F}$ が存在し、その内側と外側で分布が異なる

仮説パターン集合 $\mathcal{F}$：観測の**構造**を定義するパターンの集合

例）**地域性**：$\mathcal{F} = \{S \subseteq [47] \mid S$ **は連結**$, |S| = \ell\}$

一般的に $\mathcal{F}$ は指数的に大きい

例えば $\ell = 3$ のとき $\{$京都, 大阪, 奈良$\} \in \mathcal{F}$



赤：人口が**増加**
白：人口が**減少**

どこが難しい？

仮説パターン集合 と **レアイベント集合** は指数的に大きい

組合せ相関検定の手順

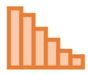
1. 観測 $x \in \{0,1\}^d$ と 仮説パターン集合 $\mathcal{F} \subseteq 2^{[d]}$ を取得
2. 観測 x の Scan 統計量 (=偏り) $K(x)$ を計算
3. 偏り $K(x)$ の P値 (=稀さ) を レアイベント集合 $\mathcal{W}$ から計算
4. P値が有意水準より低ければ帰無仮説を棄却

レアイベント集合 $\mathcal{W}$：現在の観測 x よりも**稀な**観測の集合

$\mathcal{W} = \{w \in \{0,1\}^d \mid K(w) \geq K(x)\}$

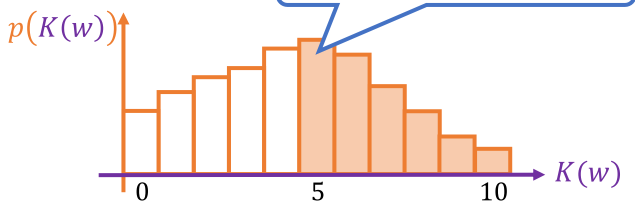
一般的に $\mathcal{W}$ は指数的に大きい

Scan 統計量：$K(x) = \max_{S \in \mathcal{F}} \sum_{i \in S} x_i$

P値：$P = \sum_{w \in \mathcal{W}} \prod_{i \in [d]} p(w_i) =$ 

例) $\ell = 10, K(x) = 5$

$K(w) = 5$ である w の確率の和



提案手法

仮説パターン集合 と **レアイベント集合** を **圧縮したまま** **P値** を計算

$$K(x) = \max_{S \in \mathcal{F}} \sum_{i \in S} x_i$$

$$\mathcal{W} = \left\{ w \in \{0,1\}^d \mid \max_{S \in \mathcal{F}} \sum_{i \in S} w_i \geq k \right\}$$

$$P = \sum_{w \in \mathcal{W}} \prod_{i \in [d]} p(w_i)$$