

レポート課題の提出方法

〆切：7月22日（金）4:00pm（厳守）

提出先：渡辺治の郵便ポストもしくは watanabe(at)is.titech.ac.jp まで PDF をメール

レポート課題

以下の各テーマごとにどれか1つを選択してレポートを作成してください。レポートは2部（2テーマ）で十分です。

テーマ# 1

- テーマ# 1の例のように、近似の見方を変えると不可能である（と思われている）問題も可能になる場合がある。そのような例を調べて紹介する。
- ストリームアルゴリズムのその他のおもしろい例を調べ紹介する。

テーマ# 2

- このテーマで紹介した飽和頻出集合列挙アルゴリズムを実際の実現し、ある程度の大きさのデータベースを用いて、その効率を検討してみよう。たとえば、Apriori アルゴリズムと比較する、あるいは、改良を入れた場合の効率向上の度合いを調べる、などがあると思う（コメント：ソースコードの提出は不要ですが、プログラミングに際して工夫した点や考えなければならなかった点、どういう実験をしたか、その結果、などの説明は必須です。）
- 逆探索法による凸多面体の端点列挙のアルゴリズムを調べ、その解説を書く。具体例などを入れてわかりやすい解説にして欲しい。計算量についても解説すること（コメント：非退化の仮定のもとでの列挙で構いません。）
- 平面上の与えられた点の集合 $P = \{v_1, \dots, v_n\}$ に対して、それを含む最小の凸多角形を P の凸包という。その凸包を P の点だけからなる三角形で分割することを P の凸包の三角形分割という。 P の凸包の三角形分割の列挙を逆探索によって行うアルゴリズムを調べ、その解説を書く。具体例などを入れてわかりやすい解説にして欲しい。計算量についても解説すること。

テーマ# 3

- このテーマで示した例のように、全体のアルゴリズムの一部分の計算は何とか望み通りに実施できると仮定して（注）、それから先はキッチリ解析し、アルゴリズムを設計している例を紹介せよ（注：もちろん、無謀に仮定しても意味がない。妥当そう

で、実際に実験してみるとうまく動くヒューリスティクスがあるのだが、それはとても複雑で解析ができない、というような場合である。)

- もしかすると、このような例はアルゴリズムの世界だけではなく数学の世界にもあるかもしれない。つまり、ある部分だけは仮定して、もしもそれが成り立つとすると ... ということ仮説の上に一つの理論が形成されているような数学の例があれば紹介して欲しい。
- 実際に AdaBoost のプログラムを作成し、実験データ (たとえば mushroom) を用いて、その性能などを実験し、その結果・解析・考察を述べよ。訓練データ (つまり事例集合) として 1000 個くらいを使い、残りのデータを使って得られた仮説の良さを評価してみるとよい。最適な仮説は何か? 訓練データを多くするとどうなるか? 高速化の工夫と効果は? 等々、いろいろと調べられると思う。

参考: データならびにサンプルプログラムは

<http://www.is.titech.ac.jp/~watanabe/class/boost/> から得られる。

テーマ # 4

- このテーマで議論した例のように、最悪時には難しいかもしれないが、ある種の入力分布を仮定したときには「平均的に」効率良く解ける問題 (あるいはそのように言われている問題) を紹介せよ。問題、分布、そして効率良いと言われている解法 (アルゴリズムまたはヒューリスティクス) についての解説である。
- 本日説明した Perturbed- k LIN Problem の最尤推定問題に対する局所探索法やメッセージ伝播法の有効性の検証。どのような場合に、どのように働くかを実験し、その結果・解析・考察を述べよ。

参考: データならびにサンプルプログラムは

<http://www.is.titech.ac.jp/~watanabe/class/maxlin/> から。

- 疎なランダム $0, 1$ -行列の最大固有ベクトルの解析。たとえば $n \times n$ のランダム行列で、確率 p で 1 , 確率 $1-p$ で 0 となるものの最大固有値, 最大固有ベクトルはどうなるか? $p = c/n$ のように小さい場合が難しい。