

大規模離散計算科学特論 (科目コードsubject code: B)

(Large-Scale Discrete Computation)

種別・単位：講義・2単位 Lecture/ 2 credits

開講期：第2学期(担当者毎の集中講義) Second semester (Intensive lectures)

担当者：鷲尾 隆 (コンピュータサイエンス専攻・大規模離散計算科学講座・内線7656)

Takashi Washio (Division of Computer Science, Ext. 7656)

白井 康之 (コンピュータサイエンス専攻・大規模離散計算科学講座・内線7656)

Yasuyuki Shirai (Division of Computer Science, Ext. 7656, shirai@erato.ist.hokudai.ac.jp)

津田 宏治 (コンピュータサイエンス専攻・大規模離散計算科学講座・内線7656)

Koji Tsuda (Division of Computer Science, Ext. 7656)

主題と目標 Subject & Objective

産業プロセスの最適化や解析、マーケティング、バイオインフォマティクスなど、様々な情報処理の局面に現れる大規模な離散構造データ(計算機が行う論理的な処理を表現したデータ)を、数学的に簡約化し効率よく計算するための各種のアルゴリズム技術とその応用について理解する。データマイニング、統計数理、機械学習、列挙アルゴリズム、簡潔データ構造、確率計算モデルなどを含む分野横断的な基盤技術とそれらを用いた最新の研究成果、および産業界への実応用の事例について学ぶ。

Algorithmic techniques for efficient manipulation of large-scale discrete structure data (description of logical procedure in computers), in various situations of information processing, such as industrial process optimization/analysis, marketing, bioinformatics, are lectured. Recent research results on fundamental and interdisciplinary techniques, including data mining, statistical mathematics, machine learning, case enumeration, succinct data structures, probabilistic computation models are shown. Industrial application examples are also presented.

授業計画 (項目, 授業実施回数, 内容) Lecture Plan

項目 Item	回数 hour	内容 Contents
離散構造・統計的因果関係のマイニングとその応用 Discrete structures, statistical causality mining and their applications	5	離散構造を対象としたデータマイニング技術の現状について概観する。また、組合せアイテム集合のマイニング技法をベースとして、グラフマイニングの技法や統計的因果推論などの統計的な理論に基づく知識発見・マイニングの技法について学ぶ。 Current techniques for data mining on discrete structures are overviewed. Various knowledge discovery and data mining methods such as itemset mining, graph mining, and statistical causal inference are lectured.
カーネル法・確率計算モデルとその応用 Kernel method, probabilistic computation models and their applications	5	サポートベクタマシンや木カーネル法, グラフカーネル等の発展的手法, またベイジアンネットワークなどの確率的計算モデルを紹介し, 生命科学分野における応用として, 膨大なデータを処理する技法や進化系統樹, 遺伝子制御ネットワークへの応用について述べる。 Support vector machine and its advanced techniques with tree/graph kernels are presented. Probabilistic computation models such as Bayesian networks are also explained. For bioinformatics applications, techniques of large-scale data processing, phylogenetic trees, and gene control networks are lectured.
頻出パターン抽出・パターン分類とその応用 Frequent pattern mining, pattern classification and their applications	5	頻出パターンや相関ルール, 系列パターンの効率的なマイニング技法や最新のパターン分類技術を概観し, インターネット上あるいは実世界における行動履歴に基づく特徴抽出手法や分析事例を紹介する。 Efficient mining/classification methods for frequent patterns, association rules and sequence patterns are overviewed. Case studies of characteristics mining/analysis from behavior records on the Internet and real world are presented.

評価・教材・受講条件等 Evaluation・Materials・Pre-requisites

《評価》 期末試験やレポート試験、授業中の質疑応答の様子等により総合的に評価する。

Evaluation will be carried out by the use of your attendance, home works and/or examinations.

《教材》 講義資料を配布する。参考書等については、適宜紹介する。

Original materials will be distributed, and references will be indicated in the class.

《受講条件等》 情報数学, データ構造, プログラミング言語, 計算理論の基礎を理解していることを前提としている。

Basic knowledge of information mathematics, data structures, programming languages, and theory of computation are needed.