HOKKAIDO UNIVERSITY

Title	ERATO湊離散構造処理系プロジェクト ： 概要紹介と最近の話題
Author(s)	湊, 真一
Description	ERATO 湊離散構造処理系プロジェクトシンポジウム（第1回） ： 第9回情報科学技術フォーラム(FIT2010) イベント企画セッション. 2010年9月8日（水）. 九州大学伊都キャンパス.
Relation	2010年度科学技術振興機構ERATO湊離散構造処理系プロジェクト講究録. p.365-369.
Issue Date	2011-06
Doc URL	https://hdl.handle.net/2115/48372
Type	conference presentation
File Information	00.IT2010-minato_slides-P.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

FIT2010イベント企画： ScienceとEngineering をつなぐ『Art』を求めて
ーERATO湊離散構造処理系プロジェクトシンポジウムー

ERATO湊離散構造処理系プロジェクト: 概要紹介と最近の話題

平成22年9月8日

研究総括　湊　真一
（北海道大学 大学院 情報科学研究科 准教授）

ERATOについて

- JSTの戦略的創造研究推進事業
  - さきがけ（牧場型）、CREST（八ヶ岳型）、**ERATO（富士山型）**
- ERATO型の特徴
  - 新しい科学技術の源流を作るような研究を支援。
  - 昭和56年発足。過去に98プロジェクトを採択。（科学技術の全分野で年4~5件）
  - プロジェクト期間：5年半　　研究費総額：10～15億円（テーマによる）
  - メンバ規模：10～15人（人件費に依存）
  - 研究総括に自由裁量を与え、分野・組織にとらわれずに機動的なプロジェクトを構成。
  - 大学への補助金ではなく、JST直轄事業。本務研究室とは別に独自のオフィスを設置。
  - ERATO専任の事務スタッフをJSTが雇用。

ERATO総括の選考過程

- 他薦およびJST独自調査により、母集団リスト（1000人規模）を構成。
- JSTで分野ごとに審査委員長（PO: プログラムオフィサー）を1名選定。POは審査員4名（うち外国人1名）を指名し、審査パネルを構成。
- 審査パネルで母集団を絞り込み、各分野で10名程度の一次候補者をノミネート。
- この段階で初めて一次候補者に通知し、研究構想の応募を依頼。
  - 候補者は突然の連絡で驚き悩む。（数名は応募辞退）
  - 研究テーマは完全に自由 （国の重点分野のどこかに該当すれば良い）（といっても、何を期待されて推薦されたのか、空気を読む必要あり）
  - 提案書は、ほぼ無地の様式で英語10ページ、和訳も添付する。
  - 研究分担者を指名・事前協議できない。個人の構想だけで勝負。
- 一次面接（または書類選考）、最終面接で、各分野1人だけ採択
  - 最終面接は英語で1時間以上
  - 審査パネルの意見を参考に、POが最終決定（合議制ではない。）

過去の情報系のERATOプロジェクト

- 北野共生システム (1998～2003)
- 今井量子計算機構 (2000～2005)
- 合原複雑数理モデル (2003～2008)
- 浅田共創知能システム (2005～2010)
- 五十嵐デザインインタフェース (2007～2012)
- 岡ノ谷情動情報 (2008～2013)

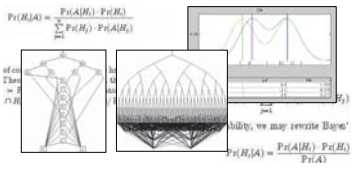
ERATO発足当初は、物理、化学、医学生命のみ（→ノーベル賞分野）情報科学系のプロジェクトが採択されるようになったのは、最近10年。しかも最初は他分野との境界領域テーマでないと取れなかった（3年前までは分野別の審査ではなく、全分野の審査員で合議だった）

アルゴリズム技術を中心とするERATOプロジェクトは画期的

本プロジェクトの概要

離散構造とその処理

- 離散構造とは
  - **離散数学および計算機科学の基礎をなす数学的構造**
  - 集合理論、記号論理、帰納的証明、グラフ理論、組合せ論、確率論などを含む



- およそ計算機が扱うあらゆる問題は、単純な基本演算を要素とする離散構造の処理に帰着される。
  - 最終的には、**膨大な数の場合分け処理**を要することが多い。
- 離散構造の処理は、計算機の様々な応用分野に共通する基盤技術
  - 典型的な応用分野： システム設計自動化、大規模システム故障解析、制約充足問題、データマイニングと知識発見、機械学習と自動分類、生命情報科学、web情報解析、etc.
  - **現代情報化社会に対する波及効果は極めて大。**

本プロジェクトの基本構想

■ 様々な工学的応用を持つ基盤技術として
「離散構造処理系」に着目し、研究開発を行う

システム設計自動化　データマイニングと知識発見
大規模システム故障解析　機械学習と自動分類
制約充足問題　生命情報科学　web情報解析

工学的応用
→ 社会への影響大

本研究の対象領域

離散構造処理系

集合理論　記号論理　帰納的証明
組合せ論　グラフ理論　確率論

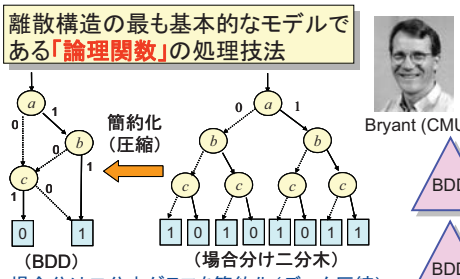
性能向上
(10倍～100倍以上)

数学的概念構造

ERATO　7

BDD（二分決定グラフ）

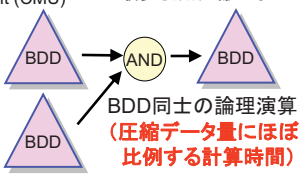
離散構造の最も基本的なモデルである「論理関数」の処理技法



・場合分け二分木グラフを簡約化（データ圧縮）
・**多くの実用的な論理データをコンパクトかつ一意に表現。**（数十～数百倍以上の圧縮率が得られる例も）

Bryant (CMU)

1986年に画期的なBDD演算アルゴリズムを提案。以後急速にBDD技術が発達。
（長期間、情報科学の全分野での最多引用文献となった）



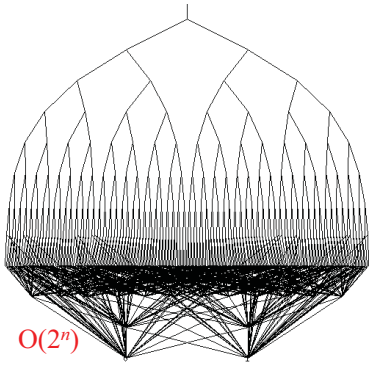
BDD同士の論理演算
（圧縮データ量にほぼ比例する計算時間）

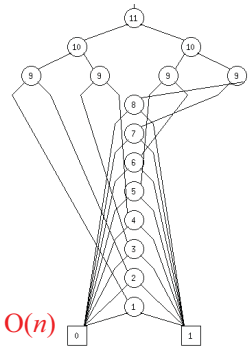
近年のPC主記憶の大規模化により、BDDの適用範囲が拡大
（特に2000年以降）

ERATO　8

BDD簡約化の効果

■ 特定の問題では、指数関数的な圧縮効果が得られる。
- 例題に依存するが、多くの実用的な問題で効果がある。


$O(2^n)$

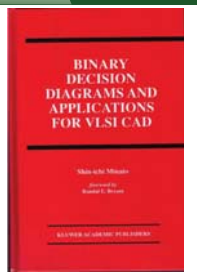

$O(n)$

ERATO　9

研究総括の本分野での業績

■ 学生時代(1988年)よりBDD研究を開始。
- 黎明期より一貫してBDD研究に従事。

■ 研究成果を単行本として米国で出版
S. Minato: “Binary Decision Diagrams and Applications for VLSI CAD” (Kluwer, 1996).
- Bryant先生に前書きを寄稿して頂いた
- 欧米の理工系大学院で副教材として使用された実績を持つ



■ 組合せ集合を表現するデータ構造「ZDD」を考案(1993).
- VLSI論理最適化システムのために開発. (1993～2003)
- データマイニング・知識処理にも応用. (2004～)
- Knuthの著名な教科書にも掲載、執筆作業に協力 (2008～2009)

ERATO　10

論理関数と組合せ集合

a	b	c	F
0	0	0	0
1	0	0	0
0	1	0	0
1	1	0	1
0	0	1	1
1	0	1	1
0	1	1	0
1	1	1	0

論理関数:
$F = (a\ b\ \sim c) \vee (\sim b\ c)$

組合せ集合:
$F = \{ab, ac, c\}$
→ ab
→ c
→ ac

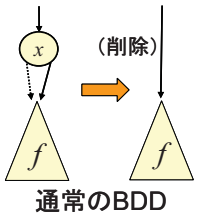
(買い物客の購入品)

■ 組合せ集合と論理関数の演算は対応関係がある。
- Union of sets → logical OR
- Intersection of sets → logical AND
- Complement set → logical NOT

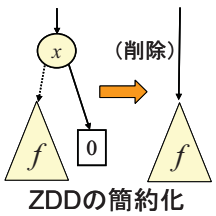
ERATO　11

ZDD（ゼロサプレス型BDD）による集合の表現

■ 「組合せ集合」を効率的に表現するためのBDDの改良
- **湊が世界で初めて考案し命名（1993年）**
- **通常と異なる簡約化規則を考案。**
- **疎な集合の族を扱う場合に著しい効果が得られる。**
（例： 商店の陳列アイテム数に比べて1顧客の購入点数は極めて少ない。）



通常のBDD



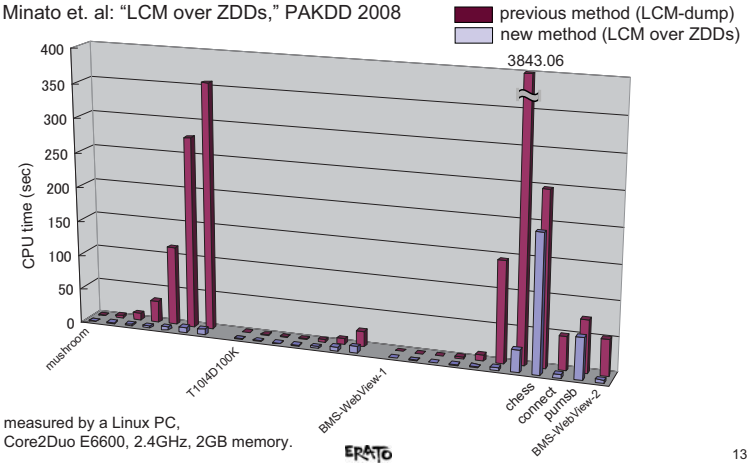
ZDDの簡約化

■ ZDDはBDDの改良技術として現在、世界的に広く使われている。
- 最近では、データマイニング分野に応用されて、画期的な有効性が示されている。（数百倍のデータ圧縮率・数十倍の処理高速化）
- 他にも応用例は増えつつある。

ERATO　12

Frequent itemset mining using ZDDs

Minato et. al: “LCM over ZDDs,” PAKDD 2008

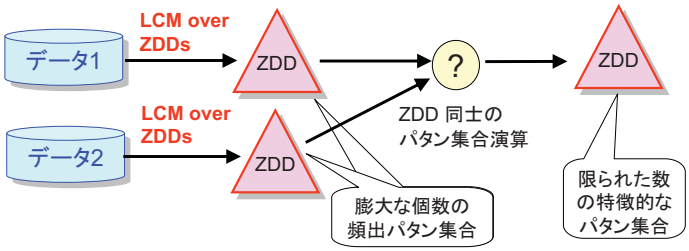


measured by a Linux PC,
Core2Duo E6600, 2.4GHz, 2GB memory.

ERATO

13

ZDDを用いたパタン集合演算



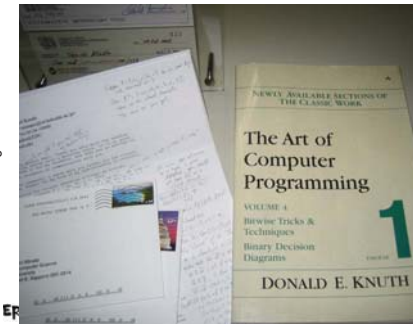
- 数十億ものパタンを含む集合を圧縮して表現し、ZDDの集合演算を使って効率よく絞込みを行える。
  - 従来の明示的な表現方法では，意味のある解析処理を現実的な時間で行うことは不可能

ERATO

14

BDD / ZDD に関する最近の話題

- Knuthの世界的名著「The Art of Computer Programming」の最新巻（Vol.4, Fascicle 1, 2009）で、BDDが取り上げられ、その中でZDDが30ページ以上に渡り詳しく解説された。
- **日本人の研究成果が、このシリーズに項目として詳細に掲載されるのは初めて。**
- Knuth氏本人から、ZDD考案者として校正作業への協力を依頼する長文のメールと手紙を受領。
- 2010年5月には湊がKnuth邸を訪問し、プロジェクトの方向性について意見交換を行った。

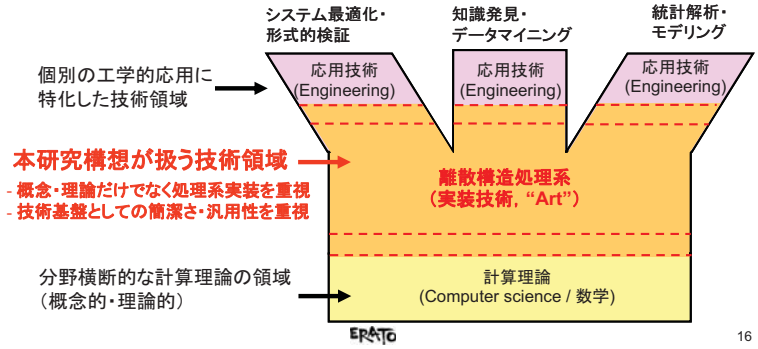


ERATO

16

本研究構想の対象領域

BDD/ZDD技術の新しい切り口として、様々な離散構造を**統合的に演算処理**する技法を体系化し、**分野横断的かつ大規模な**実問題を高速に処理するための技術基盤を構築する。



ERATO

16

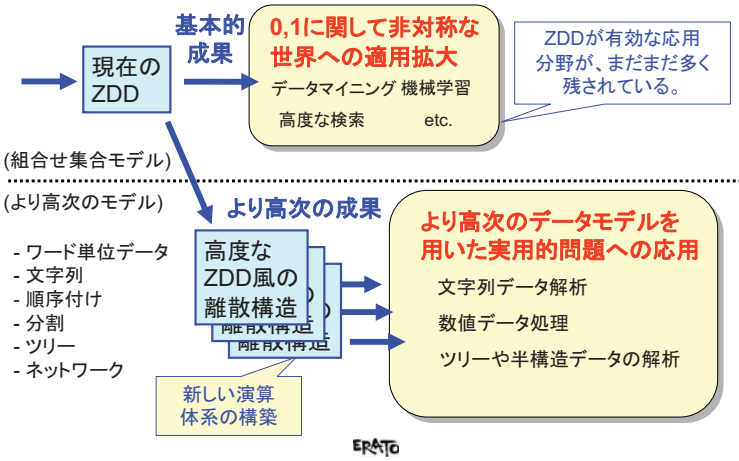
本プロジェクトの成果目標

- BDD/ZDD研究を新しい視点でとらえ直し、汎用的な離散構造処理系として再構築を行う
  - **社会的重要性が高い複数の応用に適用し、従来技法に比べて、10倍～100倍以上の性能向上を達成する。**
- 長期に渡って活用される基盤ソフトウェアを提供する。および、それを使いこなせる人材を育成し、産業界に還元する。
  - 形態素解析システムChasen, 連想検索エンジンGeta のように、**10～15年が経過しても使われる基盤ソフトウェア**
    （→ 直接的・間接的に、過去の国家プロジェクトの成果が形になったもの）
- 離散構造処理系という技術分野を深く追究し、長期に渡って国内外から参照されるような学術専門書にまとめること
  - **教科書が持つ長期的な影響力は極めて大きい。**

ERATO

17

本研究プロジェクトの技術面のポイント



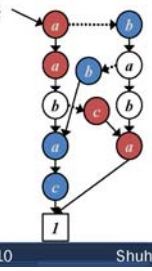
ERATO

18

SeqBDD (Sequence BDD)

- Loekito, Bailey and Pei, 2009
  - Variant of ZDD
  - 0-edges are ordered (variable order is fixed)
  - 1-edges are not ordered
  - A letter is allowed to occur multiple times in a path

{*aabac*, *aaca*, *baba*, *bbac*}



(0-terminal node is omitted)

プロジェクトの研究体制

研究グループ体制

- 離散構造基盤グループ （GL：研究総括兼任）
  - 札幌メインオフィス（北大・情報科学研究科）
  - ポスドク研究員　現在日本人5名（白井・櫻井・吉仲・川原・齋藤）
    （若手ポスドクは原則2年で次のポストを獲得するよう奨励）
  - 今後、企業出向者・海外ポスドクの勧誘を本格化
  - 事務スタッフ3名
    - 技術参事（白井、ポスドク兼務）、事務参事（齊藤）、事務員（渡辺）
- 機械学習・制約充足応用グループ（GL：産総研・津田宏治氏）
  - 東京・大岡山（東工大GCOE「計算世界観」と共同研究）
  - ポスドク研究員1名（田部井）　＋東工大・東京地区研究者と連携
- 統計・マイニング応用グループ （GL: 阪大・鷲尾隆先生）
  - 大阪・梅田・民間オフィス（阪急梅田駅北口すぐ）
  - 技術員1名（中元）補助員1名（金光）　＋関西地区研究者と連携

ERATOオフィス設置場所

- 札幌にメインオフィスを設置
  - 北大キャンパス内に設置
  - 空港・駅からの交通至便
  - 北大の関連研究者とも交流が容易
    → **研究に集中できる優れた立地環境**
- 東京・関西地区にサブオフィスを設置
  - **東工大と大阪梅田に設置**
  - **福岡（九大）にも連絡拠点（共同研究）を計画中**
  - 高品質TV会議システムを導入し、十分な意思疎通を図る。

フルタイムの若手研究者は極力、札幌に常駐させて、活性化・相乗効果を図る。

札幌オフィス

- 情報科学研究科に隣接する校舎内（工学研究科）
  - 約300m²の独立スペース（建物の端・廊下突き当り）
  - 北大情報系の研究者との交流が容易。（工学部食堂のほぼ真上の3階）
  - 耐震補強工事のための玉突き移動により、このタイミングで偶然出現したまとまったスペース
  - 部局長の特別の措置でERATOに貸出し

情報科学研究科（本務研究室）

ERATOオフィス予定地（C301~306）

当初の候補地域

北大オフィス
レイアウトイメージ

東京地区

早稲田大
林泰弘先生

電通大
植野真臣先生

東工大
佐藤泰介先生
渡辺治先生
岡本吉央先生

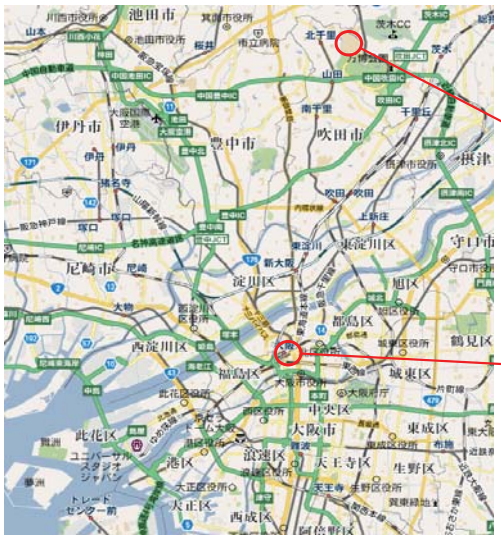
埼玉大
堀山貴史先生

NII
宇野毅明先生

産総研
津田宏治氏

25

関西地区



関学大
羽室行信先生

立命大
山下茂先生

京大
加藤直樹先生
山本章博先生他

阪大
鷲尾隆先生
浜口清治先生

梅田オフィス

大阪府大
森田裕之先生

26

ERATOセミナ

- ほぼ毎週、第一線で活躍中の研究者に講師を依頼してERATOセミナ（研究会）を開催
  - 興味のある研究者をERATOセミナ会員として登録
  - ERATOメーリングリストへの登録・イベント案内の送付
  - 定例のERATOセミナへの自由参加
  - 好きな時間に北大オフィスまたは各地区サテラボに立ち寄り、ERATOメンバとお茶を飲みながら議論して頂く
- ERATOメンバとの具体的な共同研究テーマが定まって来たら、数人単位で「ERATOミニプロジェクト」を登録
  - 連携研究者として共著論文執筆や基盤ソフト開発等に参加。
  - 必要な研究費（旅費・謝金・物品等）はJST直轄予算からサポート。
  - 知財処理等が必要な場合は別個の共同研究契約に発展させる

ERATO

27

イベント開催状況

- 5/28(金)～29(土)　キックオフシンポジウム開催
  - 全国より約40名の著名研究者が北大に集結
- 6/18（金）西尾章治郎先生　北大オフィス来訪
  - 学内関係者向けプロジェクト説明会・見学会を開催
- 6/30（水）関西キックオフセミナ
  - 約25名の関西地区研究者が参加
- 7/15（木）東京キックオフセミナ
  - 約15名の東京地区研究者が参加
- 7/30（金）OR学会NEO研究会・ERATO共催（北大）
  - 幹事：　京大・伊藤大雄先生
- 本日　FIT-2010 イベント企画（九州大学）
  - ポスター討論による研究者交流イベント

ERATO

28

今後のイベント予定

- 今年度後半（時期未定）で「ERATO合宿」を検討中
  - 北海道内の宿泊研修施設（セミナハウス・温泉ホテル等）？
- 11/18,19　国際会議 AMBN2010 共催
  （開催地：東京田町）
  - International workshop on Advanced Methodologies for Bayesian Networks
  - 海外からの著名招待講演者を計画
- 3/2(水) 情処全国大会（東工大）　企画イベント
  - AL研究会との共同企画
  - 研究者同士のポスター討論会を予定
- 3/28～30 国際会議IBISML/LLLL/DMSS 2011 共催
  （開催地：大阪中之島）
  - 著名研究者による招待講演を予定

ERATO

29

まとめ

- ソフトウェア系のERATOは、特殊な実験装置はないので、優れた研究者集団を形成するための環境作りが最も重要。
  - 日本における研究者の人材集積地は東京と関西であるが、そこから離れた札幌で実施するということが、非常に特徴的である。
  - 西尾章治郎先生（プログラムオフィサー）からは「札幌を、優秀な人材が集まる梁山泊のようにして欲しい」との激励を頂いている。
  - 東京・関西のサブグループ拠点と活発な研究交流を行いながら、優秀な若手人材を見出し、札幌に引っ張って来て、高いレベルの「梁山泊」を形成し、一定期間じっくりと研究を行う。
  - コアとなる「Art」を形成し，学界・産業界に還流させていきたい。

ERATO

30