

双方向CTLによるJava最適化器の生成

東京工業大学 数理・計算科学専攻

方 玲 佐々 政孝

概論

- CTLによる最適化

コンパイラの最適化は、プログラムによるものがほとんどだが、近年CTLという論理による最適化の研究も行われている。

それは条件付き書換え規則「 $I \rightarrow I' \text{ if } \phi$ 」を用いて表現する。

★条件付き書換え規則「 $I \rightarrow I' \text{ if } \phi$ 」を以後**最適化記述**と呼ぶ。

- この手法のメリット

- 論理式で一行か、多くても十数行で最適化を記述できる。
- 証明できる。

• 本研究の動機

- 2001年, 従来の研究として, Laceyらの研究は, 過去時制を含む CTL-FV を提唱し, 最適化器の正しさを証明した. しかし, この研究は理論に留まる.
- 2003年, 山岡らの研究は, 既存のモデル検査器SMVを用い, Laceyらの理論の一部を実装した. 過去時制を用いることができず, 無用命令除去しか扱えない.
- 2004年, 番らの研究は過去時制を除去することによってコピー伝播を行えるが, 除去処理の時間と, 除去によって式が長くなるため, 最適化時間がかかなり長い.

CTL等の論理による手法で通常のアルゴリズムと同じように最適化できるかは未解決の問題である.

• 本研究の貢献

- 通常最適化器に大幅に近づいたCTLによるJava最適化器を実装した。従来研究と比べて記述しやすい、最適化時間が格段に短い、最適化効果も十分見込める。
- 実現に当たり種々試行錯誤を行った結果、問題点が明らかになり、今後役に立つ多くの知見が得られた。

• 本研究の独創性

- モデル検査器は従来研究と異なり、何も変換せずに直接処理できる形で実装し、効率を大幅に向上した。
- 最適化記述について、従来研究は条件式が特定の番号の命令文しか対象とできないが、本研究の条件式は命令文の集合を対象とすることができる。
- 通常最適化器に近い性能を持った、CTLによる最適化器の実装は本研究がはじめてである。

時相論理と分類

- **時相論理**(temporal logic)は命題の真理値が時間とともにどのように変わるかを記述する論理体系.

「時刻0の世界」, 「時刻1の世界」, 「時刻nの世界」..

時間の概念は「時間」と「空間」

- **分類**

- 分岐的か線形的か.
- 過去時制か未来時制か.

....

LTL, CTL, CTL*, CTL_b, PCTL, NCTL

- **CTL-FV**

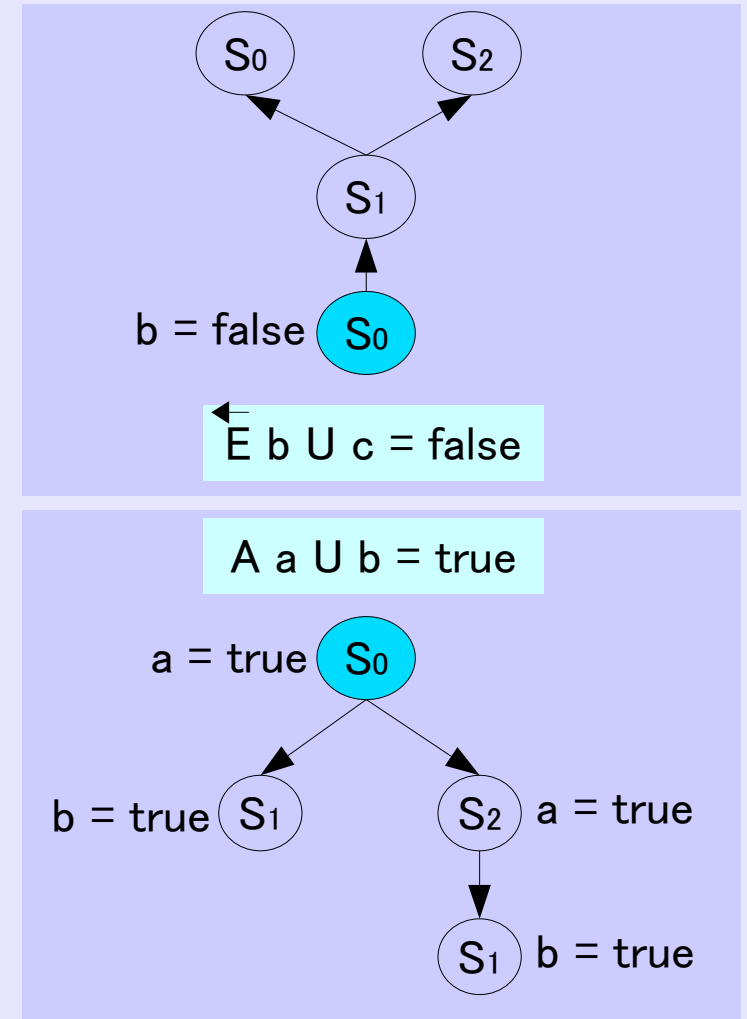
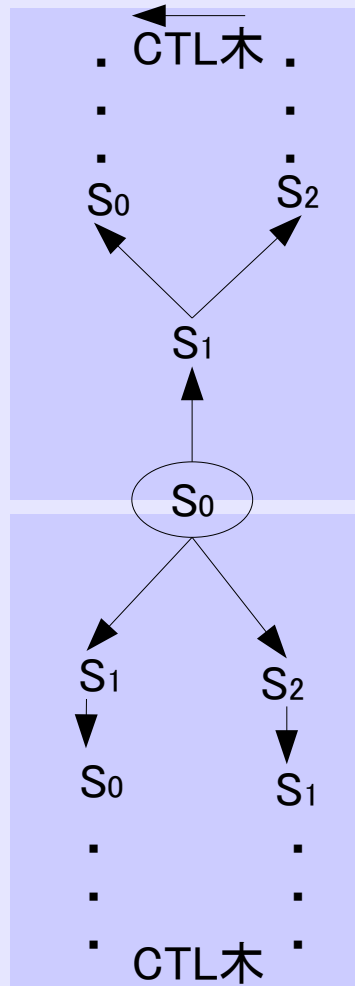
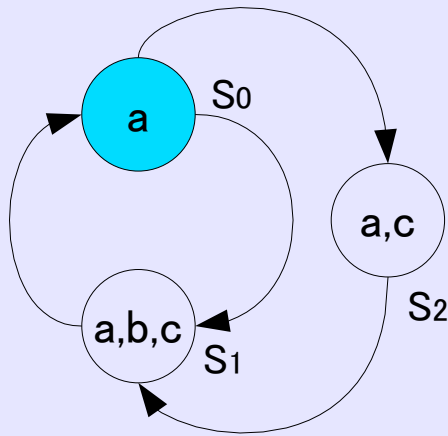
述語 ϕ の変数が特定のプログラムの変数や式の定義域を指定し, 束縛することによって関係付けられる.

CTLb (Computational Tree Logic) 演算子

経路限定子: A (futureAll), E (futureExist), \overleftarrow{A} (pastAll), \overleftarrow{E} (pastExist)

時相演算子: U (Until), X (neXt), G (Global), F (Future)

コンパイラ最適化は CTLb を用いると自然かつ簡潔に記述できる.



Kripke構造:

- S : 状態の集合
- S_{start} : 始状態の集合
- $R \subseteq S \times S$: 遷移関係
- $L: S \rightarrow 2^{Prop}$

Kripke構造

双方向CTL木

注: CTL構文木と違う

将来時制と過去時制計算の例

本研究の最適化記述

- **条件付書き換え規則:**

MATCHの形の文に対して,

CONDITIONの条件を満たした文であれば,

PROCESSの処理をする.

という流れ.

- **文法**

MATCH

変数:=式

CONDITION

point_文字列:CTL式(条件式)

point_文字列:CTL式(部分式)

edge_文字列:point_文字列 → point_文字列

PROCESS

point_文字列: Command 命令文

point_文字列: Replace 式 → 式

edge_文字列: EdgeSplit 命令文

例:

MATCH

$v := b$ (変数 := 二項式)

CONDITION

point_comp: $\text{use}(b) \wedge \text{trans}(b)$

point_spavin: . . .

point_spantout: . . .

point_insert: $\text{point_comp} \wedge \neg \text{point_spavin} \wedge \text{point_spantout}$

point_replace: . . .

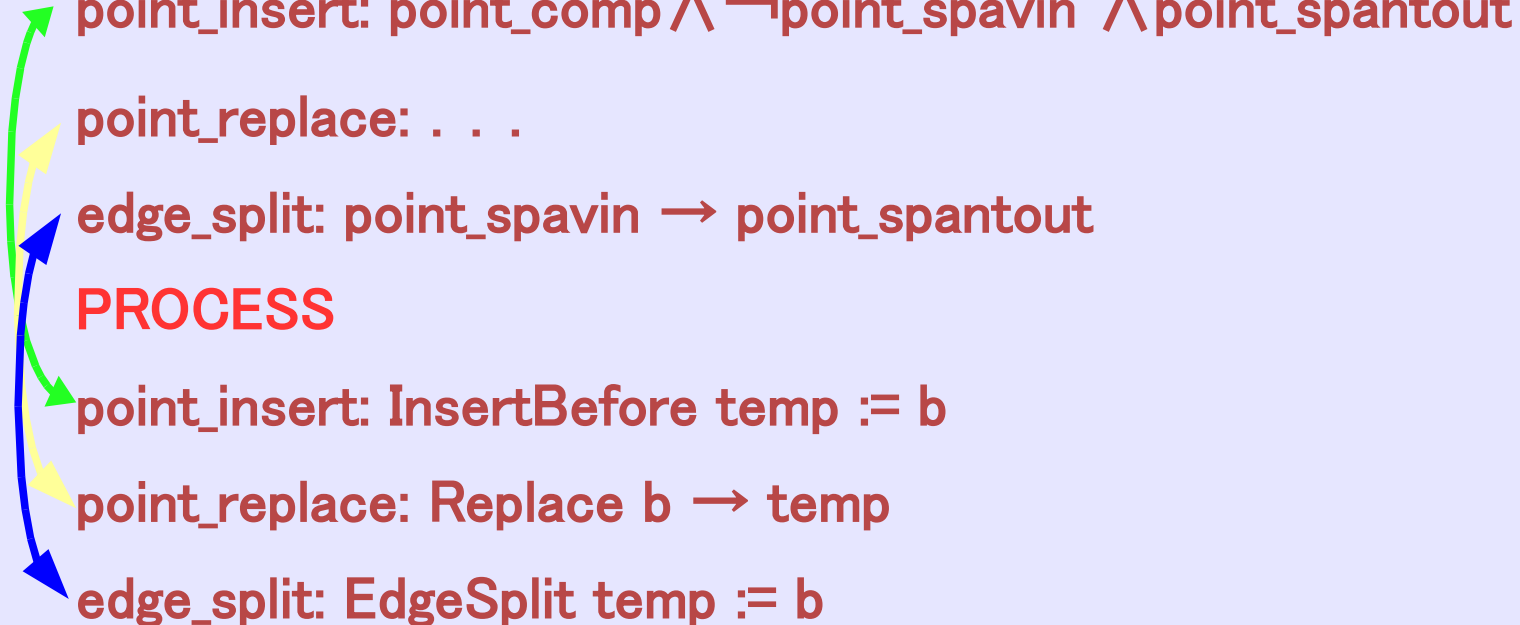
edge_split: $\text{point_spavin} \rightarrow \text{point_spantout}$

PROCESS

point_insert: $\text{InsertBefore temp} := b$

point_replace: $\text{Replace } b \rightarrow \text{temp}$

edge_split: $\text{EdgeSplit temp} := b$



- 改善点

本研究の記述

delete v := e

if point_delete : AX ((AG \neg use(v)) \vee A \neg use(v) U def(v))

既存研究記述:

n: v := e \rightarrow skip

if n \vdash AX ((AG \neg use(v))) \vee A \neg use(v) U def(v))

- モデル検査器が計算するのは特定の番号の命令文ではなく、命令文の集合である。多数の命令文を対象とした記述が容易になる。

- 最適化器の効率を向上できた。

最適化の CTLb による定式化

- 最適化の条件をもとに定式化

無用命令除去:

$$AX ((AG \neg use(v)) \vee A \neg use(v) \cup def(v))$$

- データフロー方程式をもとに定式化

$$DEAD(B) = \{x \mid x \notin \bigcup_{S \in succ(B)} LIVE(S)\}$$

$$LIVE(B) = USE(B) \cup \left[\bigcup_{S \in succ(B)} LIVE(S) \right] - KILL(B)$$

$$USE(B) = \dots$$

$$KILL(B) = \dots$$

部分冗長性除去のデータフロー方程式 (Paleriらの手法)

$$\begin{aligned}
 TRANSP_i &= trans(e) \\
 COMP_i &= use(e) \cdot TRANSP_i(e) \\
 ANTLOC_i &= use(e) \cdot TRANSP_i(e) \\
 AVIN_i &= \begin{cases} false & \text{if entry} \\ \prod_{j \in preds(i)} AVOUT_j & \text{if otherwise} \end{cases} \\
 AVOUT_i &= COMP_i + AVIN_i \cdot TRANSP_i \\
 ANTOUT_i &= \begin{cases} false & \text{if exit} \\ \prod_{j \in preds(i)} ANTIN_j & \text{if otherwise} \end{cases} \\
 ANTIN_i &= ANTLOC_i + ANTOUT_i \cdot TRANSP_i \\
 SAFEIN_i &= AVIN_i + ANTIN_i \\
 SAFEOUT_i &= AVOUT_i + ANTOUT_i \\
 SPAVIN_i &= \begin{cases} false & \text{if } i = \text{entry or } \neg SAFEIN_i \\ \sum_{j \in preds(i)} SPAVOUT_j & \text{if otherwise} \end{cases} \\
 SPAVOUT_i &= \begin{cases} false & \text{if } i = \text{entry or } \neg SAFEOUT_i \\ COMP_i + SPAVIN_i \cdot TRANSP_i & \text{if otherwise} \end{cases} \\
 SPANTOUT_i &= \begin{cases} false & \text{if } i = \text{entry or } \neg SAFEOUT_i \\ \sum_{j \in preds(i)} SPANTIN_j & \text{if otherwise} \end{cases} \\
 SPANTIN_i &= \begin{cases} false & \text{if } i = \text{entry or } \neg SAFEIN_i \\ ANTLOC_i + SPANTOUT_i \cdot TRANSP_i & \text{if otherwise} \end{cases} \\
 INSERT_i &= COMP_i \cdot \neg SPAVIN_i \cdot SPANTOUT_i \\
 INSERT_{i,j} &= \neg SPAVOUT_i \cdot SPAVIN_j \cdot SPANTIN_j \\
 REPLACE_i &= ANTLOC_i \cdot SPAVIN_i + COMP_i \cdot SPANTOUT_i
 \end{aligned}$$

部分冗長性除去のCTL記述 (Paleriらの手法)

MATCH

$v := e$

CONDITION

$\text{point_comp} : \text{use}(e) \wedge \text{trans}(e)$

$\text{point_avin} : AX (A \text{trans}(e) \cup \text{use}(e))$

$\text{point_avout} : \text{point_comp} \vee (\text{point_avin} \wedge \text{trans}(e))$

$\text{point_antout} : AX(A \text{trans}(e) \cup \text{use}(e))$

...

$\text{point_edge2} : \neg \text{point_spavout}$

$\text{edge_split} : \text{point_edge1} \rightarrow \text{point_edge2}$

PROCESS

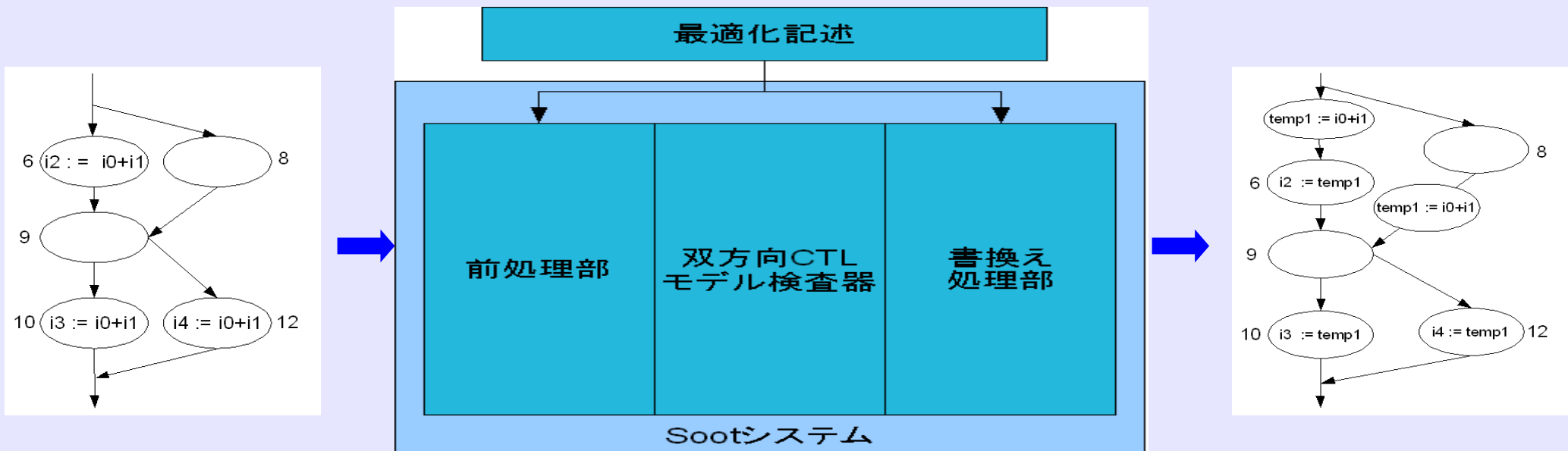
$\text{point_insert} : \text{InsertBefore temp} := e$

$\text{edge_split} : \text{EdgeSplit temp} := e$

$\text{point_replace} : \text{replace } e \rightarrow \text{temp}$

CTL_bによる最適化器

- 前処理部、モデル検査器、書換え部の3部分から構成される。



- 計算量

$$N \doteq O(n_1^{n_2} \times n_1 \times n_3) = O(n_1^{(n_2+1)} \times n_3)$$

n_1 : プログラムの大きさ

n_2 : 自由変数の数

n_3 : CTL_b構文木のノード数

実験

- テストコード

SPECjvm98の中の7つと奥村のJavaコード使って, 実験を行った.

- 実験環境

- CPU: Celeron 2GHz

- Memory: 512MB

- Soot: version 2.2.0

- 計測のオプション:

- Xint -Xms128m -Xmx128m(JITとメモリの影響を取除く)

- 適用した最適化

部分冗長性除去(共通部分式除去とループ不変式), コピー伝播(定数伝播), 無用命令除去.

(比較対照のSootが適用した最適化: 共通部分式除去, 部分冗長性除去, コピー伝播, 定数伝播&畳込み, 条件分岐の畳込み, 無用命令除去, 無用分岐除去, 無用変数除去)

最適化の処理時間

- 左図はSPECjvm98の最適化時間, 右図は奥村のコードの最適化時間(単位:秒)

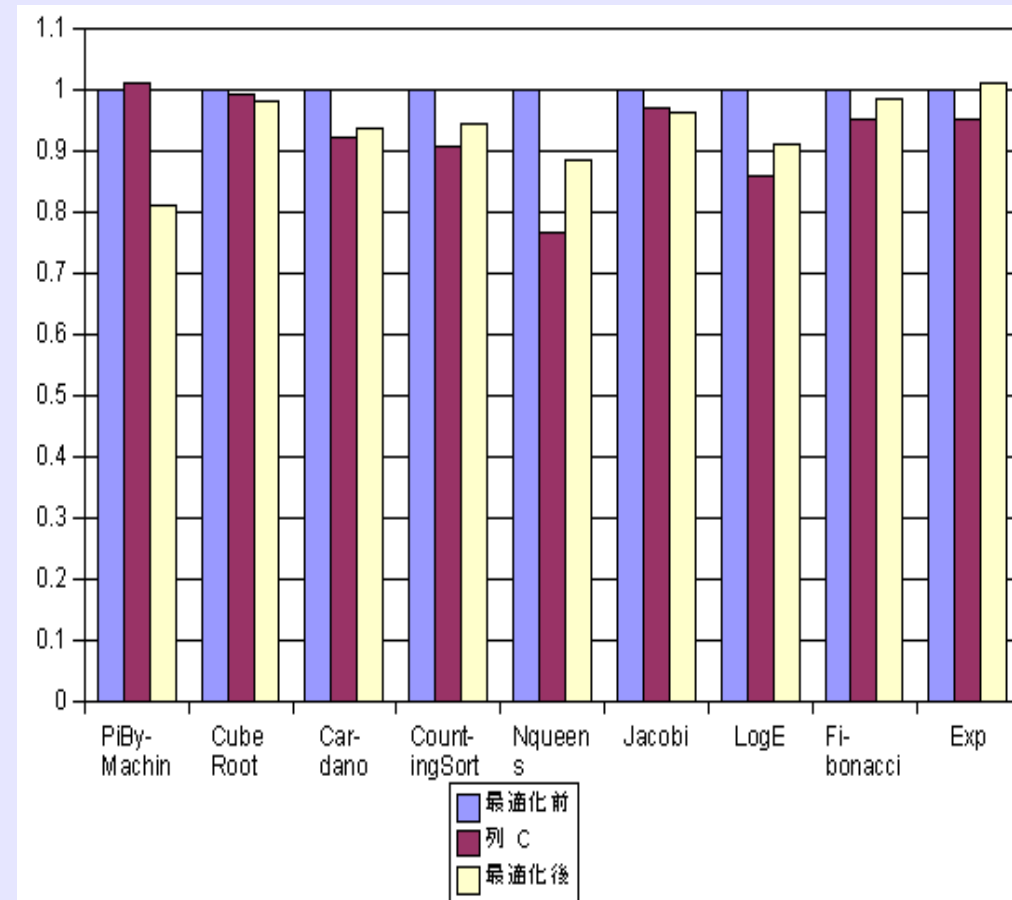
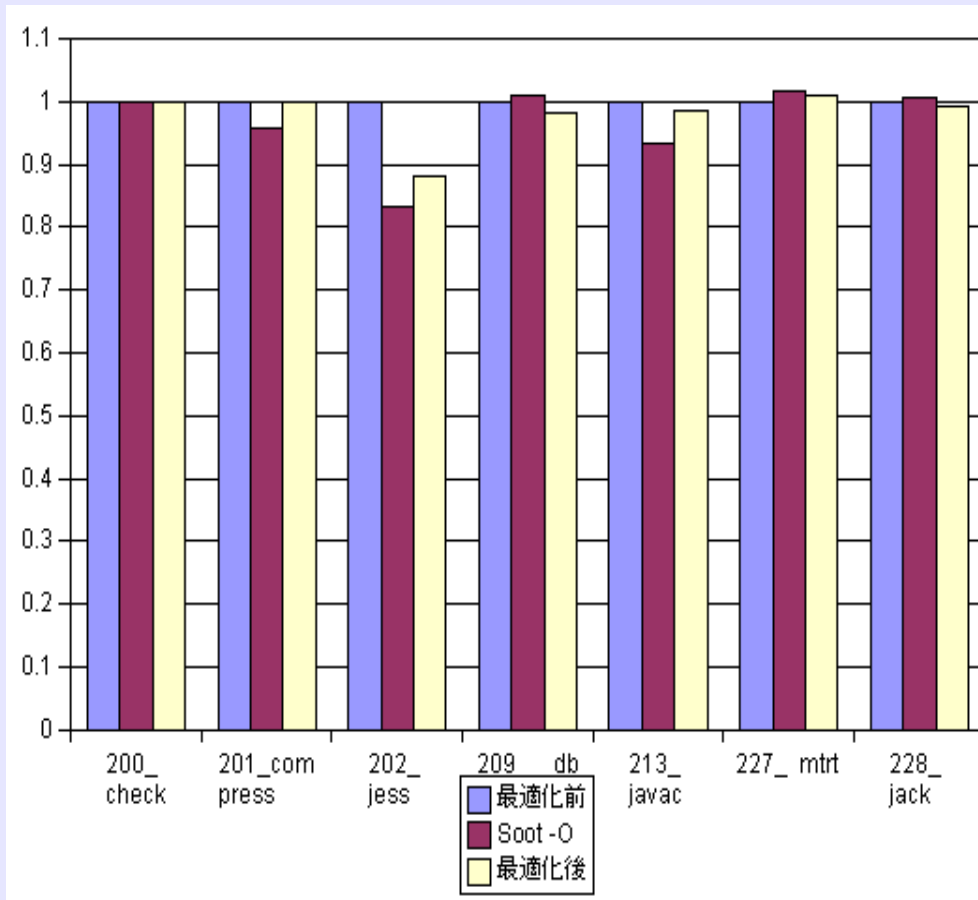
test code	行数	compile time	test code名	行数	compile time
200_check	5145	20	PiByMachin	108	0.8
201_compress	5050	29	CubeRoot	224	1.5
202_jess	26674	123	Cardano	242	7.1
209_db	5316	15	CountingSort	175	2.5
213_javac	57937	219	Nqueens	247	3.7
227_mtrt	9713	160	Jacobi	1061	48.6
228_jack	21895	316	LogE	1496	20.4
			Fibonacci	206	1.4
			Exp	1045	12.1

- データの分析:

時相論理による最適化は全探索などの処理のため, 実用的な時間内でできないと言われたが, 本研究の結果は遅くても5分以内に収まった.

最適化前後の実行時間の比較

- 左図はSPECjvm98, 右図は奥村のコード(最適化前を1に).

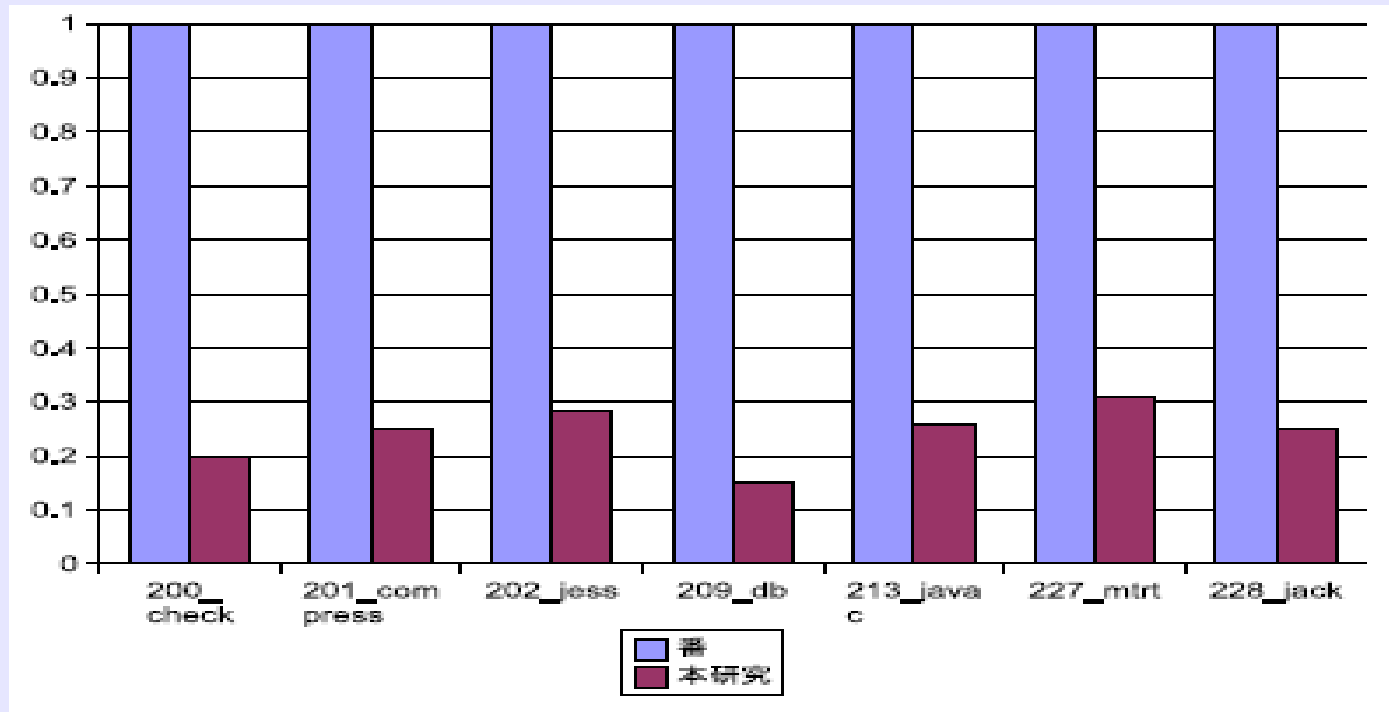


- 本研究の最適化はSootの最適化の一部だが, それなりの効果がある
- 本研究の最適化はSootの最適化の一部だが, Sootに勝ったものもある

従来手法との比較(最適化について)

- Laceyらは最適化時間や実行時間のデータを与えていない。
- 番らの手法との比較(番らの手法による最適化を1に)

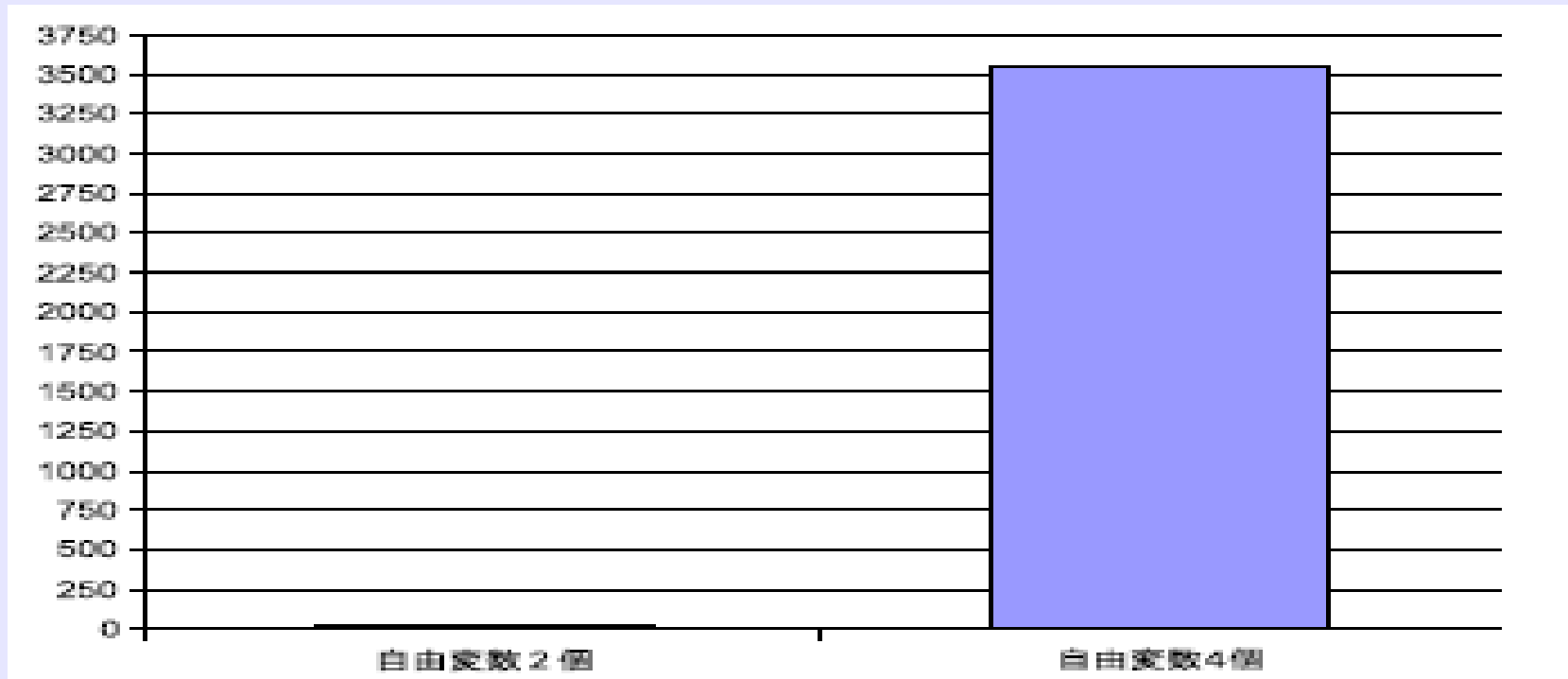
コピー伝播を適用した最適化時間を比較した。本研究の最適化時間は15%から30%になっている。



自由変数による爆発

- 図の左は自由変数2つ, 最適化時間が**20秒**, 図の右は自由変数4つ, 最適化時間が**59分**である.

本研究は実用的な最適化器は自由変数を使わないように記述すべきことを明らかにした.



考察

• CTLの表現力

細かい処理をCTL式で書こうとすると、式が長くなるし、式の正しさの証明も難しい。たとえばプロフィールを用いた部分冗長性除去、条件付定数伝播など。

• 最適化器の効率

- 自由変数の全組み合わせとモデル検査器は全探索のため、遅い

★ 最も影響するのは自由変数

• 最適化器の効果

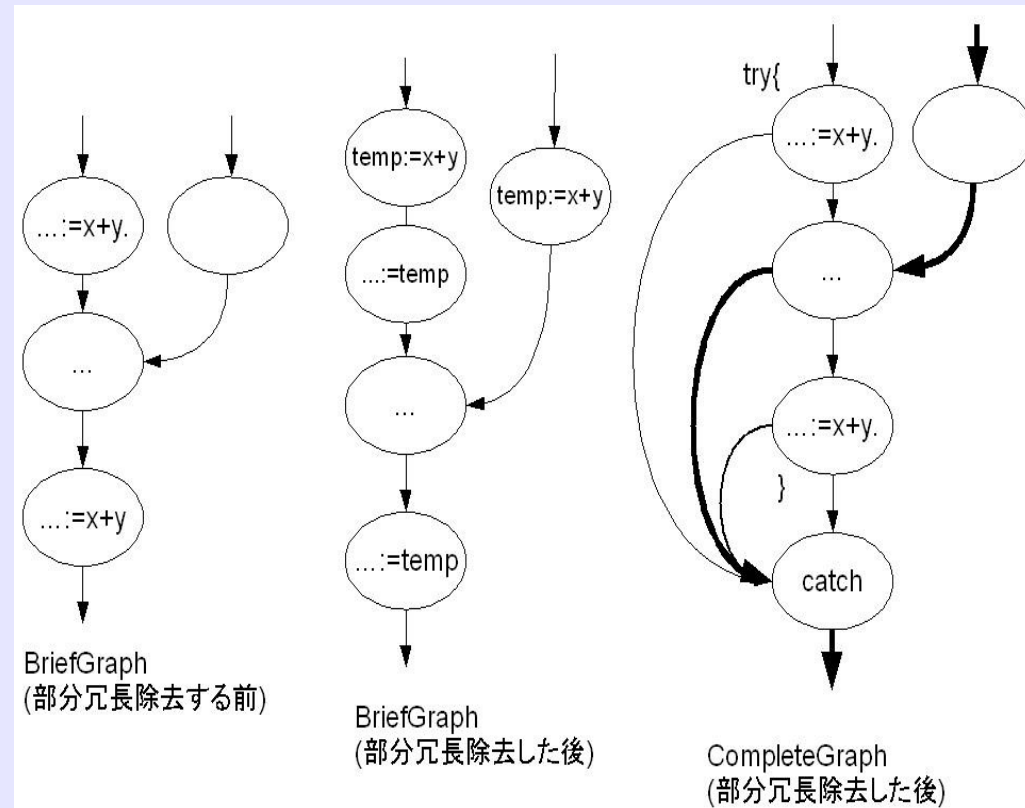
- Javaの例外を起し得る演算

と例外を超える移動は対象外

- 記述できない最適化は対象外

★ 最も影響するのは例外による

障害



本研究の位置づけ

- CTLによる最適化器の性能は、記述能力、最適化時間、最適化効果の三つの基準で評価できる。

	記述能力	最適化効率	最適化効果	特徴
Laceyら	○	×	△	最初提案、証明したが、実際の最適化の一部しか扱えない、記述は通常アルゴリズムに頼る。
山岡ら	×	×	×	過去演算子がなく、無用命令除去しか扱えない。
番ら	×	×	×	過去演算子を変換して未来演算子のみにするが、除去後の式が長い。無用命令文とコピー伝播しか扱えない。
本研究	○	○	△	モデル検査器が速い、記述しやすい。実用的な時間内で処理可能。

今後の課題

- **最適化時間の短縮**

- モデル検査器を早くするために、BDDを導入したり、部分評価などの手法で、モデル検査器の全探索を避けることが考えられる。

- ★ 自由変数は計算時間を爆発させるため、なくすべき

- **最適化の効果の向上**

- ★ 例外による最適化の阻害を克服

- ループ、for文、goto文など細かい解析。

- 条件付定数伝播など複雑な最適化をどう書くか。

まとめ

- CTL_b モデル検査器と改善した最適化記述を用いるJava最適化器を開発し、複雑な最適化を実現した。通常アルゴリズムによる最適化器の性能に大幅に近づいた。
- 最適化の時間や最適化の効果について、ベンチマークやテストコードを使ってデータを取った。CTLによる最適化の問題点をいくつか明らかにすることができた。
- 今後は、問題点の解決、性能の改善をし、CTL式に基づく実用的なJava最適化器を伝統的な最適化器の一部として組み込めるよう努めたい。