

「データ圧縮とBDDの関連性」 と自己紹介など

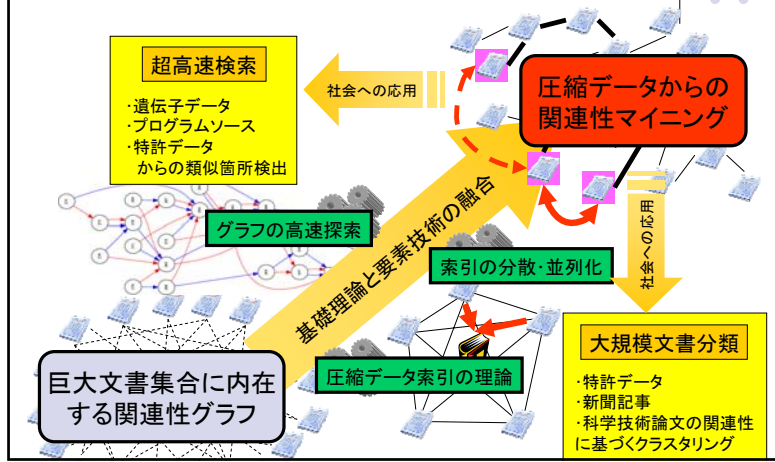
坂本 比呂志

九州工業大学 大学院情報工学研究院
科学技術振興機構 さきがけ



第1回 離散構造処理系シンポジウム, 2010-5-29

最近の研究テーマ: データ圧縮とその応用



データ圧縮の理論

Grammar-based compression algorithms

Algorithm [reference]	Approx. Ratio upper/lower	Space upper/lower	Time upper
Sakamoto[IEICE Trans. '09]	$\log^* n \log n / -$	$g \log g / \log n$	$n \log^2 n$
Charikar[IEEE Trans. '05]	$\log n / -$	n/n	n
Rytter[Theor. Comput. Sci. '03]	$\log n / -$	n/n	n
Sakamoto[J. Discrete Algo. '05]	$\log n / -$	n/n	n
Welch[IEEE Comput.]	$(\frac{n}{\log n})^2 / \frac{3n}{\log n}$	n/n	n
Larsson[...]			
Nevill-M...[...]			
Kieffer[...]			
Apostolico[DCC '98]	$(\frac{n}{\log n})^2 / 1.37$	$n \log n / n$	$n \log^2 n$

Grammar-Based Compression

→ 精度保証

← 効率化の技法

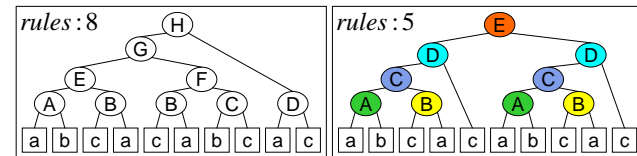
BDD

注釈: $\log^* n = \log$ がとれる回数, $n \leq 2^{65536}$ のとき $\log^* n \leq 5$

データ圧縮

Grammar-Based Compression (文法圧縮)の概念

- ただひとつの文字列を生成するCFG
 - $A \rightarrow \alpha | \beta$ の規則を禁止 (ORの禁止)
 - 変数のループを禁止 ($A \rightarrow^* \alpha A \beta$ の禁止)



- 圧縮サイズ = CFGサイズ (生成規則の右辺の総和長)
 - より単純に変数の種類 |N| で置き換え可能

生成規則木(PRT)による圧縮索引

丸山君(竹田研), 馬場君(定兼・山下研)との共同研究



投稿中のため割愛

5

実験結果



投稿中のため割愛

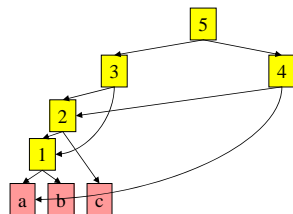
6

データ圧縮とBDDの関連性

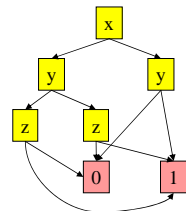
■ 構造が似ている

- どちらも有向非巡回グラフ(DAG)
- 根からたどった先に何らかの答えが書いてある

CFG圧縮のDAG表現



BDDのDAG表現



7

データ圧縮とBDDの関連性

■ 最適解に対する計算複雑さが似ている

- どちらもNP-hard

■ 近似困難性が似ている

- BDDは定数倍近似不可能
- CFGは(おそらく)不可能

■ 近似可能性が似ている？

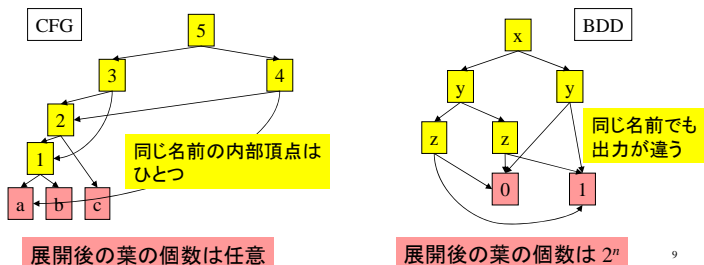
- CFGはlog n-近似可能
- BDDは(log n-近似可能)？

ここを示したい

8

データ圧縮とBDDの関連性

- 一方を他方の概念に還元できないか？
 - 互いの分野の技法が利用できる
 - 二つのギャップ (1)葉の個数, (2)内部頂点の意味

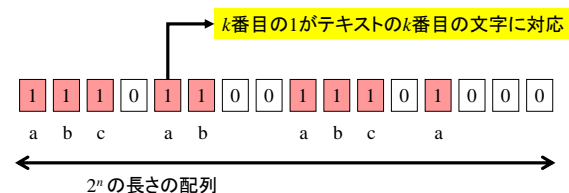


データ圧縮とBDDの関連性

- 葉の個数の対応付け
 - テキスト長 k を超える最小のべき数 2^n を考える

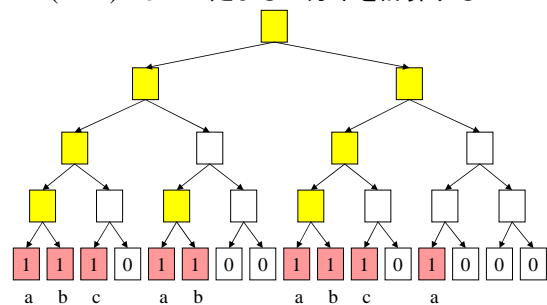
$k = |abcababca| = 9$ の場合は $2^n = 16$

2^n 個のうち k bit に 1 を (適当に) 割り振る



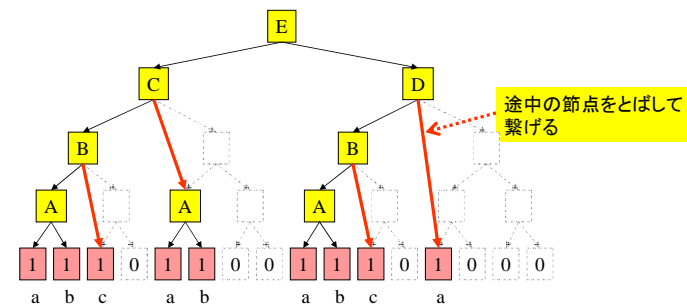
データ圧縮とBDDの関連性

- 葉の個数の対応付け
 - 2^n の葉を持つ全二分木上で、隣接する 1 の最近共通祖先 (LCA) によって定まる二分木を計算する



データ圧縮とBDDの関連性

- 葉の個数の対応付け
 - 得られる二分木はCFGのある構文木とちょうど対応する



データ圧縮とBDDの関連性



■ 残された問題

BDDの変数順序がCFGの何に対応しているか？

