

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 29 日現在

機関番号：62615

研究種目：若手研究(B)

研究期間：2015～2016

課題番号：15K16057

研究課題名(和文) 問題の構造を利用した高速並列SATソルバの研究開発

研究課題名(英文) Development of Fast Parallel SAT Solvers Using Problem Structure

研究代表者

園部 知大 (Sonobe, Tomohiro)

国立情報学研究所・ビッグデータ数理国際研究センター・特任研究員

研究者番号：50747269

交付決定額(研究期間全体)：(直接経費) 1,000,000円

研究成果の概要(和文)：充足可能性問題(SAT問題)は効率的に解くのが難しいと信じられている問題であるが、ソフトウェアや論理回路の検証等を行う際に使用される。そのため、この問題を解くSATソルバの高速化の意義は高く、本研究課題では大規模な問題を処理するための高速並列SATソルバを、グラフ理論の手法を用いて研究開発を行った。作成した並列ソルバは、SATソルバの性能を競う国際的な競技会で上位に入賞を果たし、その性能の高さを示した。

研究成果の概要(英文)：The Boolean satisfiability problem (SAT problem) is considered to be difficult to solve in an efficient manner. Because SAT problems are derived from real-world applications such as software and circuit verification, development of fast SAT solvers for solving SAT problems is important. In this research, we have developed parallel SAT solvers by using methods in graph theory. The developed solvers won prizes in the international SAT solver competitions.

研究分野：SATソルバ

キーワード：充足可能性問題 SATソルバ 探索 並列 グラフ

1. 研究開始当初の背景

充足可能性問題(SAT 問題)は最も基本的な NP 完全問題として知られており、効率的に解を発見するのが難しいと信じられている問題の一つである。この問題の解を探索するソルバは SAT ソルバと呼ばれ、回路検証やソフトウェア検証等の、現実のアプリケーションの基盤として使用されており高速化の意義は高い。

SAT 問題は真偽どちらかの値を取りうる Boolean 変数から構成される、節と呼ばれる論理式を用いて構成されており、全ての節を真とするような変数割り当ての有無を問う問題である。多くの SAT ソルバはバックトラック型の探索アルゴリズムに基づき、探索中に発見される矛盾から新たな節を学習していくことで効率的に探索を進める。近年の主流な並列 SAT ソルバは、この逐次型アルゴリズムに基づくソルバをワーカーとして、探索空間の分割を行わずに、全ワーカーを競争的かつ協力的に動作させるポートフォリオ戦略に基づくものが多い。

ポートフォリオ戦略に基づく並列 SAT ソルバの強みは、分割統治型の並列ソルバにおける煩雑な探索空間の分割を必要としない点にあり、あるワーカーが解を発見した時点で他のワーカーの探索結果を待たずして探索を打ち切ることができる。ポートフォリオ戦略では、ワーカーがそれぞれ異なる探索を行うことが前提で、それによってワーカーどうしの探索の重複を避け多様な探索を実現する。この、ワーカーの探索を可能な限り異なるものにするには多様化(diversification)と呼ばれ、主に各ワーカーの探索パラメータを変更することで実現されてきた。しかし、パラメータを変更してもソルバの探索挙動は完全には予測できないため、ポートフォリオ型戦略では並列数が増加してワーカーの数が増えると、ワーカー間の探索がより重複しやすくなる欠点が考えられてきた。なお、各ワーカーは独立に探索を行いつつも、他のワーカーと学習節を共有することで協調的な探索も行う。

研究代表者はこれまで高速 SAT ソルバの研究開発に従事し、逐次ソルバと並列ソルバの開発を行ってきた。逐次ソルバでは、探索の頑健性を向上させるためのリスタート戦略に関して新しい手法の提案を行い、SAT ソルバの性能を競う国際的な競技会で上位に入賞しその性能の高さを示した。その手法をポートフォリオ型並列ソルバに応用し、既存の並列ソルバの改良に成功した。それらの成果から、実アプリケーションの問題から派生する SAT 問題に備わる、特定の変数間に存在する強い依存関係(問題の構造)を利用することで SAT ソルバの性能の改善が可能であるという着想を得た。特にポートフォリオ型並列 SAT ソルバにおける既存の多様化手法は単純なパラメータの変更に基づいており、より探索の重複を避ける強い多様化を実現

するためには問題の構造を考慮する必要がある。また多様化に加えて、ワーカー間で共有される学習節に関しても問題の構造を取り入れることが可能である。

2. 研究の目的

本研究の目的は、実アプリケーションから派生した SAT 問題に内在する問題の構造を利用し、高速な探索を実現することである。特に、近年研究が盛んな並列 SAT ソルバの高速化に焦点を当て、大規模な問題の解の高速な探索を目指す。SAT 問題の構造に関しては、近年、問題をグラフとしてとらえ、密に関係する変数集合が抽出可能であることが示されている。この変数集合を探索時に抽出することで、並列 SAT ソルバの探索に活用する。

具体的には、

- (a) 抽出した複数の変数集合を並列ソルバの各ワーカーに割り当て、各ワーカーは割り当てられた変数集合に対する集中的な探索を行うことで、ワーカー間の探索空間の重複を削減する
- (b) 抽出した変数集合を用いて、探索の任意の時点において、どの集合に対する探索が集中的に行われているかを判定し、他のワーカーから共有される学習節の選別を行うことで、ワーカー間の通信量の削減をする

これらの手法をもとに並列 SAT ソルバを実装し、毎年開催される国際的な SAT ソルバのための競技会である SAT Competition や SAT Race で上位への入賞を目指す。

3. 研究の方法

現実の問題から派生する SAT 問題は数十万変数から構成されるような巨大な問題も存在するが、それらの問題の解探索が実用的な処理時間で可能な理由の一つとして、問題に内在する構造の存在が挙げられている。現実の問題はランダムに生成された問題と比べて、特定の変数に対する値が決定すると多くの変数の値も決定するような変数間の強い論理的関係があり、この特定の変数間の強い依存関係が問題の構造として考えられている。

近年、問題の構造を定量的にとらえるために、グラフ理論の観点から SAT 問題を解析し、変数をノード、同一制約内の関係をエッジとして表現したグラフに対して、エッジが密に連結しあう変数集合(コミュニティ)の抽出が行われてきた。コミュニティ内ではエッジが密に、コミュニティ間ではエッジが疎に連結されているほど良い分割として考えられ、その良さを測る指標としてモジュラリティという値が考案されている。モジュラリティを最適化するようなコミュニティ分割を行うと、現実の問題から派生する SAT 問題のグラフは高いモジュラリティ値を示すことが実験的に示されている。つまり、変数

の総数が多い問題でも、論理的な依存関係は特定の変数間(コミュニティ内)だけで完結している可能性が高い。このコミュニティ構造を探索の開始前に抽出することで、事前に問題の構造を把握することが可能となり、探索の指針として利用ができる

本研究では、巨大な Web グラフやソーシャルグラフに対して高速かつ高いモジュラリティ値でコミュニティを抽出可能な Louvain 法を SAT ソルバ内部に実装し、入力となる SAT 問題をグラフに変換してコミュニティの抽出を行う。ベースとなるポートフォリオ型並列 SAT ソルバは、近年の代表的なソルバである MiniSAT をベースとして開発が進められた PeneLoPe を使用する。実装に関しては研究代表者のこれまでの開発経験を活かして探索効率を低下させないように手法の追加を行う。

4. 研究成果

SAT 問題のコミュニティ構造を利用する並列 SAT ソルバの開発および、コミュニティ構造をもとにした SAT ソルバの解析を通して、以下の研究開発成果を得た。

(1) コミュニティ構造を利用した並列 SAT ソルバにおける節共有手法の提案

ポートフォリオ型の並列 SAT ソルバの学習節の共有に関して、全ての学習節を共有するのではなく、探索に役立つ長さの短いものや Literal Block Distance (LBD)の低いものが共有の対象になる。しかし、SAT ソルバの探索では多くの節を学習し、すべての節を保存しておくともメモリの消費が顕著になり、その管理コストが探索の妨げになるため、定期的に役に立たない学習節を削除する機能が組み込まれている。そのため、他のワーカーから輸入した学習節が探索に影響を与えること無く削除されることが多々観測される。本質的に、それらの学習節を共有することは該当するワーカーの管理コストが増大するだけであるため、その共有を避けることで処理性能の向上が期待できる。

一つのコミュニティを構成する変数集合は互いに密な関係にあり、SAT ソルバの探索は個々のコミュニティ内の変数に値を割り当てて探索を進めて行くと考えられる。本提案では学習節を構成する変数が所属するコミュニティを元に、各ワーカーの直近の探索状況から共有がのぞましい学習節を割り出し、対象の学習節が探索に影響を与えらると推測されるワーカーにのみ共有を行う。

具体的には、各ワーカーは現時点で探索対象としてよく選ばれている変数を含むコミュニティを共有が望ましいコミュニティとして登録し、学習節を他のワーカーから輸入する際に、登録したコミュニティに含まれる変数を含む学習節のみに限定する。これにより、現時点の探索に最も関連する学習節のみを共有可能となる。例えば学習節 (x y

z) があり、 x がコミュニティの ID が 0、 y が 2、 z が 3 にそれぞれ属するのであれば、コミュニティ 0、2、3 を登録しているワーカーにのみこの学習節は共有される。なお、学習節が加わることでグラフの構造が変化していくため、定期的にコミュニティ構造の抽出をやり直して、各ワーカーは登録するコミュニティ情報の更新を行う。

本提案手法を代表的な並列 SAT ソルバの一つである PeneLoPe に実装し比較実験を行った。ベンチマークは SAT Competition 2014 のアプリケーション部門の 300 問を使用し、一問あたりの実行処理時間の上限は 5000 秒とした。スレッド数は 8 である。5000 秒以内に解けた問題数は、元の PeneLoPe は解の存在する充足可能な問題が 112 問で解の存在しない充足不可能な問題が 116 問、提案手法を実装した PeneLoPe は充足可能な問題が 118 問で充足不可能な問題が 115 問であった。充足不可能な問題に対しては大きな効果は確認できなかったが、充足可能な問題に対してはある程度の性能向上を達成した。

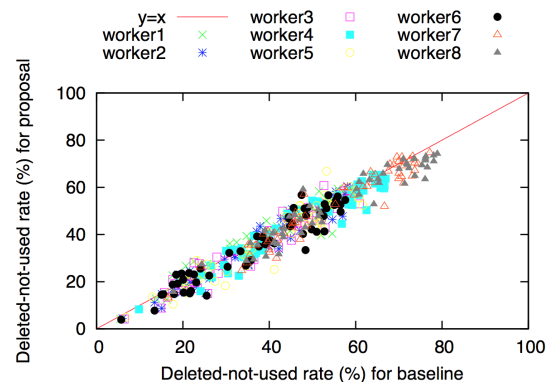


図 1 充足可能な問題に対する未使用で削除された学習節の割合

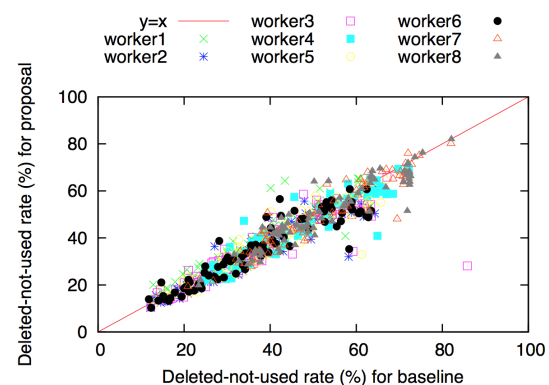


図 2 充足不可能な問題に対する未使用で削除された学習節の割合

図 1 と 2 は、充足可能/不可能な問題に対して、共有したものの一度も探索で使われずに削除された学習節の割合をワーカー毎にプロットした図であり、各点が各問題に対応する。x と y 軸は提案手法の有無による割合を示しており、提案手法により未使用のまま

削除される学習節の割合が全体的に減少していることを示している。この実験ではワーカー数(スレッド数)が 8 と比較的少ないが、よりワーカー数が多い環境や、通信コストの大きい分散環境では通信量が増大するため、本提案手法による性能向上が期待できる。

(2) 学習節の評価手法の改善

SAT ソルバの探索では多くの節が学習され、その数は指数的に増えていく。学習された節は探索空間の枝刈りに役立つものの、学習された全ての節を保存していると、メモリの容量を圧迫し、その管理コストが増大してしまう。そのため、SAT ソルバは定期的の一部の学習節の削除を実行している。特に近年のソルバは積極的な削除を行っており、一定期間毎に保存している学習節を半分にまで削減するものもある。

削除を行う際に、SAT ソルバは保存している学習節を探索の役に立つ順にソートする。この役に立つ度合いを測る指標として、最初は学習節の長さ(短いほど探索空間の剪定を促す)を基準にしていたが、近年のソルバは Literal Block Distance (LBD) を使用しているものが多い。LBD は、対象の節内の Literal Block の総数を表し、Literal Block は値が代入された変数とそれに付随して強制的に値が決定した変数の集合である。直感的には、同一 Literal Block 内の変数は互いに依存関係が強く、一つの変数の値が定まると他の全ての値が決定する可能性が高いため、それに伴い対象の節はブロックの個数に応じた変数の値が定まると節の値が決定する可能性も高い。つまり、LBD の値が対象の節の実質的な長さを表していると考えられており、LBD の低い節ほど探索の役に立つとみなされる。

LBD の特徴として、対象の節を再評価した際に、その値が変化する点が挙げられる。これは、新たに別の学習節が加わることで、Literal Block の形が変化するためである。これによって LBD の値が低くなり、対象の節が持つより本質的な LBD の値に更新されていく。そのため、本来 LBD の低い学習節でも、削除時に高い値であると削除すべきでなくても対象となる可能性が高い。つまり、本質的に低い LBD を持つ学習節を見つけ出して保存することで、探索の効率化が実現できる。

本研究では、LBD の値に加えて、将来 LBD の更新が発生する度合いを数値化するための指標 Block Deviation (BD) を提案した。BD は、対象の節の中の Literal Block の大きさの偏差を値としたものであり、この値の大きさが LBD の更新に関わると考えた。

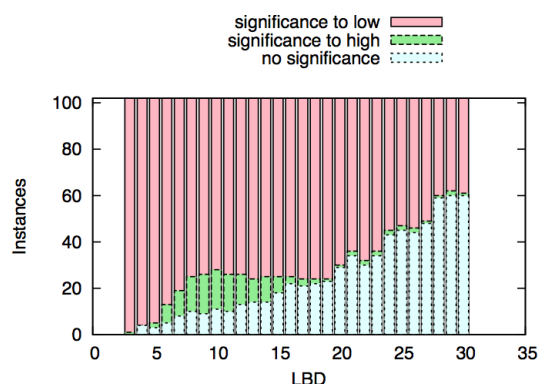


図 3 充足不可能な問題に対する Block Deviation の検定結果

図 3 は BD の値の大小が LBD の更新にどのように関連しているかを統計的に検定した結果である。SAT Competition 2014 で使用された充足不可能な問題(うち探索が終了した 102 問)に対して、各 LBD において BD の値が十分に小さい時に更新が生じたものがピンク色で、BD の値が大きいに更新が生じたものが緑色で、大きくも小さくもないと判定されたものが水色で示されている。全体的に、特に探索の使用頻度が高い LBD の低い学習節において、ピンク色の占める面積が大きいため、BD の値が小さいほど LBD の更新が発生しやすいと考えられる。この結果を実際に学習節の削除に応用したところ、より短い処理時間で多くの問題を解くことに成功した。

本提案は並列 SAT ソルバにも応用が可能であるため、将来的には並列ソルバのワーカーへの応用および、学習節の共有における基準としての使用を検討している。

(3) 並列 SAT ソルバの国際的な競技会への参加

研究代表者は、SAT 問題のコミュニティ構造を利用して、各ワーカーの探索を異なるコミュニティに属する変数へと促す community branching というポートフォリオ型並列 SAT ソルバの研究開発を行ってきた。この成果を実装した並列ソルバ CBPeneLoPe を、SAT ソルバの性能を競う国際的な競技会である SAT Race 2015 の並列 SAT 部門に出場させ、総合 4 位(著者別 3 位)に入賞を果たした。

さらに、研究成果(1) コミュニティ構造を利用した並列 SAT ソルバにおける節共有手法の提案、を加えた新たな CBPeneLoPe と、機械学習の手法を取り入れた新たなソルバ CCSPeneLoPe を翌年度の競技会 SAT Competition 2016 に参加させ、それぞれ並列ソルバ SAT 問題部門において 3 位と 2 位に入賞を果たした。これらのソルバの詳細は SAT Race 2015 (<http://baldur.iti.kit.edu/sat-race-2015/>) と SAT Competition 2016 (<http://baldur.iti.kit.edu/sat-competition-2016/>) のページに description が公開されている。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計1件)

(1) Tomohiro Sonobe: Looking Inside Literal Blocks: Towards Mining More Promising Learnt Clauses in SAT Solving, The 28th International Conference on Tools with Artificial Intelligence (ICTAI 2016), pp. 972-979, San Jose, USA, (November 2016). (査読有)

〔学会発表〕(計1件)

(1) 菌部知大: コミュニティ構造を利用した並列 SAT ソルバの学習節共有手法の提案. 第31回人工知能学会全国大会, 北九州小倉, 2016年6月.

6. 研究組織

(1) 研究代表者

菌部 知大 (SONOBE, Tomohiro)

国立情報学研究所・JST ERATO 河原林巨大
グラフプロジェクト・特任研究員

研究者番号: 50747269