1E3-OS-4-3

SAT ソルバの学習節に対する新しい評価手法の提案

Proposal of new evaluation methods for learned clauses in SAT solver

奥川 巧[*1] Takumi Okugawa　安本 猛[*1] Takeru Yasumoto　越村 三幸[*2] Miyuki Koshimura　藤田 博[*2] Hiroshi Fujita　長谷川 隆三[*2] Ryuzo Hasegawa

[*1]九州大学大学院システム情報科学府 Graduate School of Information Science and Electrical Engineering

[*2]九州大学大学院システム情報科学研究院 Faculty of Information Science and Electrical Engineering

SAT problem is a fundamental problem in computer science and has many application areas such as planning, scheduling and hardware verification. Therefore it is important to make a faster solver. In this paper, we propose True LBD which is a new criteria of learned clauses based on LBD. Furthermore, we propose new three methods for updating the LBD value. Finally, we experimentally compare them.

1. はじめに

充足可能性問題 (Boolean Satisfialbility Problem : SAT 問題) とは,「与えられた論理式が真となるような変数割り当てが存在するか否かを判定する」問題である．SAT 問題は NP 完全問題 [2] であることが知られており，回路検証やネットワーク検証，スケジューリングなどの計算機科学における様々な問題が SAT に還元可能である．それゆえに，SAT 問題は古くから重要視され，研究されている．

SAT 問題を解くプログラムを SAT ソルバと呼ぶ．奇数年に SAT Competitions[3], 偶数年に SAT-Race[4] という SAT ソルバの性能を競う協議会が開催されており，近年の SAT ソルバは以前に比べ高速化している．

本研究では，現在様々な SAT ソルバで学習節評価基準として広く使われている Literals Blocks Distance (LBD)[5] の問題点を示し，それを修正した新評価基準 True LBD を提案する．また，LBD の値の更新についても新たに 3 つの手法を提案する．最後に，LBD と LBD 値更新手法の最適な組み合わせについて，調査した結果を述べる．

2. Literals Blocks Distance

Literals Blocks Distance(LBD) とは，現在様々な SAT ソルバで使用されている学習節評価基準である．これは SAT 2009 Competition で SAT+UNSAT 部門で 2 位，UNSAT 部門で 1 位を獲得した SAT ソルバ glucose[6] が初めて実装したものである．LBD は，節の各リテラルがどのレベルで割り当てられたかという情報を基に節を評価する．ここで，同じレベルで割り当てられたリテラル群をリテラルブロックと定義すると，LBD は以下のように定義される．

定義 ある節 C が n 個のリテラルブロックで構成されているとき，節 C の LBD の値は n である．

LBD の値が小さい節ほど重要度は高くなる．これは，同一レベルで割り当てられているリテラル同士が互いに関連性を持っていることを利用している．つまり，LBD の値が小さければ，少ない変数決定で節が伝播する可能性が高いということになる．

また，LBD の決定は節中全ての変数が割り当てられている場合にのみ可能である．推論中に SAT ソルバがその状態を検知できるのは，節で単位伝播が起こった時と節が矛盾節となった時のみであるので，LBD の値はその時にのみ変動する．

図 1 に LBD のイメージを示す．図中の英字を含む正方形は，節が含むリテラルを表している．また，[ ] 内部の数字は，その変数が割り当てられてた時点の決定レベルを表す．同じ決定レベルで割り当てられたリテラルは同じ色になっている．

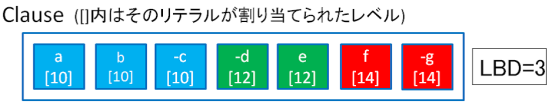

図 1: LBD のイメージ

一般的に，節が短い (節の含むリテラル数が少ない) ほど重要度が高いと考えられている．短い節の LBD の値は，確かに小さくなりその考え方に沿っている．だが，長い節の中にも LBD の値が小さいものが存在する．

例えば，長い節であっても LBD の値が 1 であれば，連鎖的に全ての変数が割り当てられ，伝播が発生したということであり，その後も同じように伝播が発生する可能性が高い．このように，節が長い場合でも LBD の値が小さければ短い節に匹敵する重要度を持つことが考えられる．

3. True LBD

学習節の評価尺度に LBD を導入しているソルバとして，MiniSat[1] をベースに開発された glucose や glueminisat[7] がある．しかし，これらのソルバに実装されている従来の LBD は，MiniSat の仕様により LBD の定義に厳密には則していない．

SAT ソルバの推論手法の基となっている DPLL アルゴリズム [8] の 1 リテラル規則 (one literal rule) では，節中にレベル 0 で割り当てられた False の値を持つリテラルは節中から削除する．しかし，MiniSat ではその仕様上，単位節となったリテ

連絡先: 奥川 巧，九州大学大学院 システム情報科学府 情報学専攻，〒 819-0395 福岡県福岡市西区大字元岡 744 番地ウエスト二号館 914 室，092-802-3599, takumi.okugawa@inf.kyushu-u.ac.jp

ラルの逆リテラルを保持する節に対して特に処理を行わない．その逆リテラルは，決定レベル 0 で割り当てられたリテラルと認識するようになっている．これは，その処理を行うことで探索速度が低下することを恐れたためだと考えられる．

しかしこの場合，レベル 0 で割り当てられた逆リテラルが存在するだけで，図 2 のようにその節が持つ LBD の値は DPLL アルゴリズムに従った場合の値よりも 1 だけ大きくなる．LBD の値は小さいほど良いので，この時の重要度は節が本来持つはずだった値よりも下がることになってしまう．

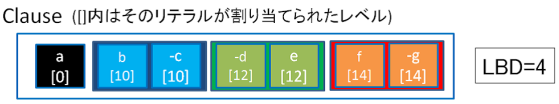

図 2: 通常の LBD の実装

本節では DPLL アルゴリズムに従った場合の LBD を利用する実装手法を提案する．その手法は，図 3 のように，レベル 0 のリテラルブロックを無視することで簡単に実現できる．

図 3: TrueLBD の実装

この手法を glucose が実装している LBD と区別するために True LBD と名付ける．また，これ以降は通常の LBD を Glue LBD と呼ぶことにする．

4. LBD の更新手法

LBD が静的な値で無いことから，LBD の実装に当たり，様々な更新手法が考えられる．本節では，LBD の更新手法を 4 つ紹介する．

- **Lowest LBD**
  LBD の値が小さくなれば値を更新し，それ以外の場合は前回の値を保持するという実装．glucose, および SAT 2011 Competition にて SAT+UNSAT 部門で 2 位，UNSAT 部門で 1 位を獲得した glueminisat もこれを採用している．
- **Initialized LBD**
  LBD の値を初期値から一切変更しない実装．
- **Newest LBD**
  LBD の値を常に最新の値に更新する実装．
- **Recent LBD**
  最近の LBD の値を優先するが，これまでの値も考慮する実装．LBD の値は次の漸化式 1 に従う．なお，OPT は実行時に与えられるパラメータで 0 から 1 までの値を取る定数である．0 であれば Initialized LBD，1 であれば Newest LBD と同一である．

$$LBD(t) = OPT \times NewLBD + (1 - OPT) \times LBD(t-1) \quad (1)$$

5. 実験

本章では，まず Glue LBD に全ての更新手法を実装し，評価を行う．その後，True LBD についても同じように評価実験を行う．

そして最後に，最も良かった手法と従来手法の比較と考察を行う．

5.1 節削除の実装方法

全ての手法は MiniSat に glucose の学習節削除戦略を加えた MiniSatAgg に対して実装している．glucose の学習節削除戦略では，式 (2) が成り立つ場合に節削除が行われる．ここで，conflicts は推論中に発生した矛盾の総数，ConflLimit は推論中に変動する閾値である．

$$conflicts \geq ConflLimit \quad (2)$$

ConflLimit は初期値として定数 BASE を持ち，節削除が行われる度に定数 INC と削除回数 numReduceDB を用いて

$$ConflLimit = BASE + INC \times numReduceDB \quad (3)$$

と更新される．本実験では定数を以下のように設定している．

- BASE = 30000
- INC = 10000

5.2 実験環境

本実験では全て以下の環境で行っている．

CPU Core Duo 3.33GHz

メモリ 8GB

OS LINUX

また，使用する問題セットは以下のものを用いた．

- SAT 2009 competition Application 部門にて使用された 292 問
- SAT 2005 competition から 176 問 (SAT05)

SAT 2009 competition で使用された 292 問は以下のように分けられている．

- SAT 2009 competition 用に作成された 165 問 (SAT09)
- SAT 2007 competition で使用された 74 問 (SAT07)
- SAT-Race 2008 で使用された 27 問 (SATRACE08)
- SAT-Race 2006 で使用された 26 問 (SATRACE06)

問題 1 つに対する実行時間は 1200 秒で，その時間内でソルバが答えを返せなかった場合は問題を解けなかったことになる．

表 1: GlueLBD の比較

	SATISFIABLE	UNSATISFIABLE	TOTAL
MiniSatAgg	234	111	345
glucose2.0	233	121	354
glueminisat2.2.5	230	123	353
InitializedLBD	241	116	357
LowestGlueLBD	236	113	349
NewestGlueLBD	241	113	354
RecentGlueLBD(0.250)	232	116	348
RecentGlueLBD(0.500)	240	114	354

5.3 GlueLBD の評価実験

GlueLBD について，それぞれの更新手法を実装した結果を表 1 に示す．Recent LBD における () 内の数値は，式 1 における OPT の値である．

表からわかる内容を以下にまとめる．

- 広く使われている Lowest LBD は Activity を利用するよりも確かに良い性能を見せるが，他と比べ性能は悪い．
- Newest LBD は，SAT である問題に対して Initialized LBD と同等の性能を誇る．そして Recent LBD(0.50) も同じように SAT である問題に対して性能が良い．

結果より，GlueLBD を用いる限り，初期値で固定する場合が最も性能が良いことがわかった．これは，更新による計算コストが大きいことが起因していると考えられる．LBD の更新は単位伝播の度に行われる．単位伝播は SAT ソルバの実行時間の 70〜90 %を占める処理であるため，当然その計算コストは大きくなる．つまり，GlueLBD では更新によって評価値の精度を上げても，更新の計算コストよりも大きな効果は得られなかったということになる．

結論として，GlueLBD を用いる限り，LBD は初期値を利用する方が更新値を利用するよりも有用である．これは，GlueLBD が DPLL に基づいた更新を行っていない弊害の可能性が高い．

5.4 TrueLBD の評価実験

TrueLBD について，それぞれの更新手法を実装した結果を表 5.4 に示す．GlueLBD の評価実験同様，RecentLBD における () 内の数値は，式 1 における OPT の値である．

表 2: TrueLBD の比較

	SATISFIABLE	UNSATISFIABLE	TOTAL
MiniSatAgg	234	111	345
glucose2.0	233	121	354
glueminisat2.2.5	230	123	353
InitializedLBD	241	116	357
LowestTrueLBD	237	113	350
NewestTrueLBD	245	115	360
RecentTrueLBD(0.250)	238	117	355
RecentTrueLBD(0.500)	242	116	358

表からわかる内容を以下にまとめる．

- GlueLBD を利用した場合と比べ，更新を行わない Initialized LBD 以外の全ての結果が向上した．
- GlueLBD の結果と同様，最新の LBD を利用することで SAT である問題に対する性能が向上している．
- Recent LBD(0.25) の結果から，最新だけでなく一連の LBD の値を考慮することは UNSAT である問題に対して効果的である．

TrueLBD を用いることで，更新における計算量を超えた更新効果を発揮することができた．また，LBD は最新の結果を用いるのが最も効果があるようである．これは，LBD の保持している伝播関係が作成時以降に学習節の削除や変数 Activity の変化で変動しているからだと予想される．

UNSAT に対しては，どの手法も glucose，glueminisat を超えることはできなかったが，提案手法の中では Recent LBD(0.25) の性能が最も良かった．UNSAT である問題が解けるということは単位節がより生成されやすいと言えるので，平均的に小さい値に更新されるような節が学習節を生成する際に役立つのではないかと予想される．

5.5 NewestTrueLBD と従来手法の比較と考察

本節では，新手法で最も性能が良かった NewestTrueLBD の性能を考察する．

以下に NewestTrueLBD，LowestGlueLBD，MiniSat の 3 つの手法において解けた問題数の時間推移を表すグラフを示す． グラフより，NewestTrueLBD は，時間が進めば進むほ

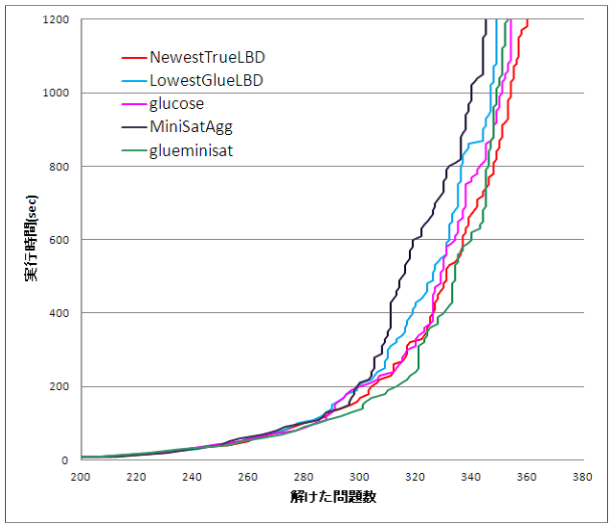

図 4: LBD の比較

ど問題を解いているのに対して，他の手法は終盤ほとんど伸びていない印象を受ける．この傾向から，実行時間が増えれば増えるほどこの差は大きくなっていくのではないかと予想される．

次に，NewestTrueLBD を用いることで，MiniSatAgg からどれだけ問題を解く速さが向上したのかに注目する．

図 7 は各問題における実行時間を比較したグラフとなっている．このグラフでは，点が斜線の上に集中しているほど横軸の手法の方が性能が高いことを示す．逆であれば縦軸の性能が高いことになる．

図をみると，実行時間が短いものを除けばほぼ全ての点が斜線よりも上にあるのがわかる．これは，NewestTrueLBD の実装により MiniSatAgg よりも解ける問題数も増えたが，ど

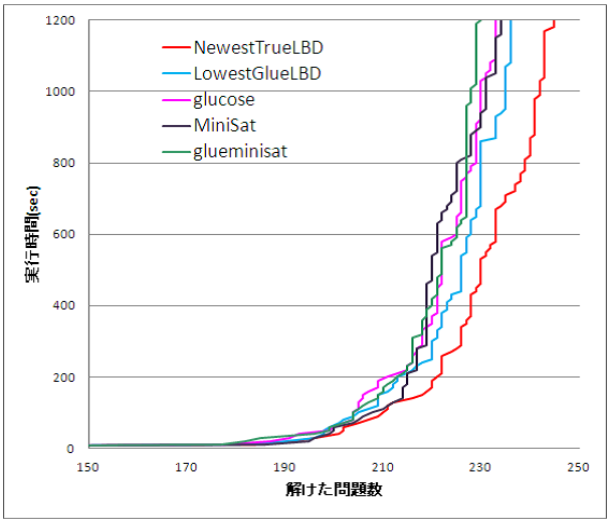

図 5: SAT 問題のみでの LBD の比較

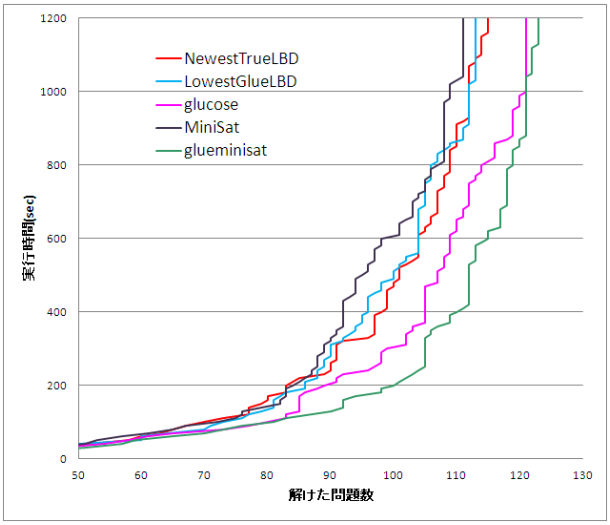

図 6: UNSAT 問題のみでの LBD の比較

ちらのソルバでも解ける問題に対しても，NewestTureLBD を用いることで解く速度が上昇していることがわかる．

結論として，NewestTrueLBD を用いることで，より多くの問題を解くことに成功した．また，この手法は時間をかければかけるほど解ける問題数が増えるのではないかと予想できる優れた手法である．更に，単位伝播毎に LBD の計算を行うという計算コストがかかっているにも関わらず MiniSatAgg よりも問題を解くまでの実行時間が減少している．これは，計算にかかるコストよりも，NewestTrueLBD の良い節を選択する効果が高かったことを示している．つまり，良い学習節を保持できているため，効率的な探索を行えている．

6. まとめ

本研究では，既存の学習節評価基準である LBD の問題点を示し，それを修正した新評価基準 TrueLBD を提案した．また，LBD の値の更新についても新たに 3 つの手法を提案し，実験によって学習節の評価基準と，LBD 値更新手法の最適な

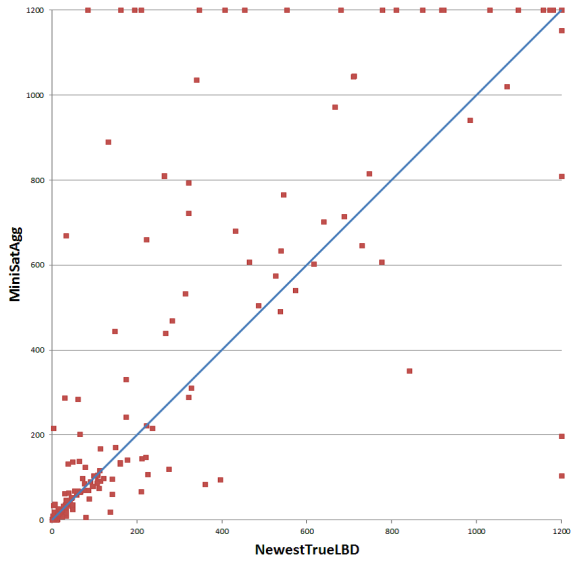

図 7: NewestTrueLBD と MiniSat の比較

組み合わせを調査した．その結果，TrueLBD と NewestLBD の組み合わせが最適であり，既存の手法よりも多くの問題が解けるようになり，問題を解く速度も向上することが判明した．

今後の課題として，LBD の更新はその方法により探索の性質が変わることから，どれか 1 つではなく複数の更新手法を組み合わせて利用する方法を考案していきたい．

謝辞 本研究は科研費 (20240003,21300054) の助成を受けたものである．

参考文献

[1] Eén, Niklas and Sörensson, Niklas : An Extensible SAT-solver, SAT 2003, pp. 502-518 (2003)

[2] Cook, Stephen A.: The complexity of theorem-proving procedures, Proc. 3rd Annual ACM Symposium on the Theory of Computing, New York, 151-158（1971）

[3] SAT Competitions.
http://www.satcompetition.org/

[4] SAT-Race.
http://baldur.iti.uka.de/sat-race-2010/

[5] Gilles Audemard and Laurent Simon.: Predicting Learnt Clauses Quality in Modern SAT Solvers, IJCAI 2009, pp. 399-404 (2009)

[6] Audemard, G. Simon, L.: GLUCOSE: a solver that predicts learnt clauses quality, SAT 2009 competitive events booklet, preliminary version, pp.7-8 (2009)

[7] 鍋島英知，岩沼宏治，井上克巳: GlueMiniSat 2.2.5：単位伝播を促す学習節の積極的獲得戦略に基づく高速 SAT ソルバー, 日本ソフトウェア科学会第 28 回大会, 6E-1 (2011)

[8] Davis, Martin and Logemann, George and Loveland, Donald: A machine program for theorem-proving, Commun. ACM, 5(7), pp.394-397 (1962)