STLS: Cutset-Driven Local Search For MPE

Alon Milchgrub
The Hebrew University of Jerusalem, Israel
alon.milchgrub@mail.huji.ac.il

Rina Dechter
University of California, Irvine
dechter@ics.uci.edu

Abstract

In this paper we present a cycle-cutset driven stochastic local search algorithm which approximates the optimum of sums of unary and binary potentials, called *Stochastic Tree Local Search* or $STLS$. We study empirically two pure variants of $STLS$ against the state-of-the art GLS^+ scheme and against a hybrid.

Introduction

The problem of optimizing discrete multivariate functions or "energy functions" described as sums of potentials on (small) subsets of variables is one of fundamental importance and interest in a wide variety of fields, such as computer vision and graphical models. In the context of the latter, conditional probability tables (CPT) are used to describe the relations between the variables of a model. Instances of this problem arise in the form of Most Probable Explanation (MPE) problems, where finding a maximum of such an energy functions composed of the CPTs, translates to finding an assignment of maximum probability given some partial assignment as evidence.

Background

Definition 1 (Energy Minimization Problem)**.** let $\bar{x} = x_1, \ldots, x_N$ be a set of variables over a finite domain $\mathcal{D}$, let $\varphi_i : \mathcal{D} \to \mathbb{R}$ for $i \in \{1, \ldots, N\}$ be unary potentials, and let $\psi_{i,j} : \mathcal{D}^2 \to \mathbb{R}$ for a subset of pairs $E \subseteq \{\{i,j\} : 1 \leq i < j \leq N\}$ be binary potentials, then *the problem of energy minimization* is finding

$$\bar{x}^* = argmin_{\bar{x}} \sum_i \varphi_i(x_i) + \sum_{\{i,j\} \in E} \psi_{i,j}(x_i, x_j)$$

Definition 2 (Cycle-Cutset)**.** Let $G = (V, E)$ be an undirected graph. A *cycle-cutset* in G is a subset C of V, such that the graph induced on $V' = V \backslash C$ is acyclic.

Given an instance of the energy minimization problem, a primal graph can be built, where every variable x_i is assigned a vertex, and two vertices x_i and x_j are connected if there exists a potential $\psi_{i,j}$. If the resulting graph is acyclic, the problem can be solved efficiently using Belief Propagation (BP) (Pearl 1982; Bertele and Brioschi 1972). If the graph is not acyclic, a cycle cutset can be generated and an optimal assignment to the forest variables given the assignment to the cutset variables can be found in a method known as "*cutset-conditioning*"(Pearl 1988; Dechter 1990).

Algorithm 1 pseudo code of STLS/STLS*.
STLS and STLS* differ in the restart and the cutset generation procerdures.

Input: Graph $G = (V, E)$ annotated with potentials φ_i and $\psi_{i,j}$.
Output: An assignment $\bar{x}$ which achieves the minimum energy found in time t.

// Identify nodes not part of any cycle and set aside
$T \leftarrow$ FindTreeVariables(G)
$V' \leftarrow V \backslash T$
$\bar{x} \leftarrow$ InitializeValues
while $runtime < t$ **do**
  $C \leftarrow$ GenerateCutset(V', E)
  $T' \leftarrow V' \backslash C$; $F \leftarrow T \cup T'$
  // Alternate BP on forest variables and local search on cutset variables
  repeat
    $\bar{x}_{prev} \leftarrow \bar{x}$
    $\bar{x}_F \leftarrow$ BeliefPropagation($G, \bar{x}, F$)
    $\bar{x}_C \leftarrow$ SubsidiaryLocalSearch($G, \bar{x}, C$)
  until $x_{prev} = x$;
  if *stagnated* **then**
    $\bar{x} \leftarrow$ InitializeValues
  end
end

STLS: Stochastic Tree-based Local Search

(Pinkas and Dechter 1995) suggested iteratively conditioning on a different cutset and finding exact optimal solution on the rest variables as a possible scheme for dealing with cycles in the graph. However, they did not go further to establish the capabilities of this method. The algorithm can additionally perform regular local search on the cutset variables. The operation of of $STLS$ is given in Algorithm 1. In every iteration an optimal assignment to the forest variables

Set		STLS		STLS*		Hybrid		GLS+ Random		GLS+	
(# inst.)		Best	Ratio	Best	Ratio	Best	Ratio	Best	Ratio	Best	% of best
Grids	mean	2 (0)	0.99	1 (1)	0.99	**18** (1)	**1.02**	0 (0)	0.75	0 (16)	95%
(21)	max	3 (1)	**1.04**	1 (7)	1.03	**16** (1)	1.03	0 (0)	0.86	0 (10)	
CSP	mean	14 (4)	**1.19**	13 (4)	1.17	13 (5)	1.17	18 (7)	1	**19** (6)	77%
(29)	max	18 (2)	**1.19**	18 (3)	**1.19**	16 (12)	1.18	**21** (4)	1	20 (4)	
Protein.	mean	1 (4)	**1**	0 (0)	0.99	4 (1)	**1**	3 (3)	**1**	**5** (1)	**100%**
(9)	max	**5** (2)	**1**	0 (1)	0.99	**5** (2)	**1**	3 (2)	**1**	4 (1)	
SGM.	mean	32 (29)	0.94	0 (5)	0.85	37 (40)	0.99	53 (33)	**1**	**57** (32)	**100%**
(90)	max	37 (32)	0.96	30 (23)	0.96	42 (43)	0.99	55 (35)	**1**	**52** (35)	

Table 1: Statistics for average and maximal results over 10 runs on sets from PIC2011. Best is the number of the instances for whom the algorithm achieved the best (second best) result. Ratio is the average ratio of the result obtained by the algorithm to that of regular GLS+. For GLS+ the average ratio of the result to the best overall result is presented.

is generated given the values of the cutset variables. Therefore, the energy of the system can not increase from iteration to iteration and the resulting algorithm is a local search algorithm finding the optimal solution on all the forest variables in every iteration.

Experiments

Three different versions of $STLS$ - $STLS$, $STLS^*$ and a $STLS$-GLS^+ hybrid - were test and compared to GLS^+ (Hutter, Hoos, and Stützle 2005), another local search algorithm and considered for many years to be the state-of-the-art. All algorithms were run 10 times on problems from the Grids, CSP, ProteinFolding and Segmentation problem sets of the PASCAL2 Probabilistic Inference Challenge (PIC2011)[1] (see (Lee, Lam, and Dechter 2013) for a summary of the statistics of these benchmark sets). The resulting energies were sampled after 0.1 seconds and after 3 minutes (1 minute for Segmentation problems), and all the results of a given instance were linearly normalized to the interval $[0, 1]$, mapping the worst result to 0 and the best to 1. In $STLS$ and $STLS^*$ the variables were initialized to an *undefined* value, thus ignored until obtaining a valid value. GLS^+ was initialized either randomly or customarily using the Mini-Bucket heuristic. The cutset variables were updated using the Hopfield Model activation function as the local search algorithm mentioned in line 10 of Algorithm 1.

As can be seen in Table 1, although the GLS^+ variants do manage to produce the best results more often, especially in the Segmentation benchmark, the average ratios of the average and maximal results obtained by the various algorithms to those of the classic GLS^+ range from slight superiority for GLS^+ on the Segmentation benchmark to significant dominance of the $STLS$ algorithms on the CSP benchmark. This implies that while GLS^+ manages to produce the best results in many case, it does not significantly outperforms $STLS$ and considerably struggles on some instances.

Conclusion

We presented an algorithm which combines the notion of cycle-cutset with the well known Belief Propagation algorithm to achieve an approximate optimum of a sum of unary and binary potentials. This is done by the previously unexplored concept of traversal from one cutset to another and updating the induced forest, thus creating a local search algorithm, whose update phase spans over many variables (the forest variables). We presented experiments suggesting this algorithm is on-par with the state-of-the-art in general and significantly outperforms it on some benchmarks.

In future work, the algorithm should be extended to handle potentials of higher arity than 2. Importantly, $STLS$ yields strong local optima, and therefore, in the limit it is as good as max-sum/min-sum belief propagation in quality, while it can be more effective computationally (i.e., guaranteed convergence). Comparing with specific loopy belief propagation scheme is left for future work as well.

[1] http://www.cs.huji.ac.il/project/PASCAL/index.php

References

Bertele, U., and Brioschi, F. 1972. *Nonserial Dynamic Programming*. Orlando, FL, USA: Academic Press, Inc.

Dechter, R. 1990. Enhancement schemes for constraint processing: Backjumping, learning, and cutset decomposition. *Artif. Intell.* 41(3):273–312.

Hutter, F.; Hoos, H. H.; and Stützle, T. 2005. Efficient stochastic local search for mpe solving. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, IJCAI'05, 169–174. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Lee, J.; Lam, W.; and Dechter, R. 2013. Benchmark on daoopt and gurobi with the pascal2 inference challenge problems. In *DISCML*.

Pearl, J. 1982. Reverend bayes on inference engines: a distributed hierarchical approach. In *in Proceedings of the National Conference on Artificial Intelligence*, 133–136.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Pinkas, G., and Dechter, R. 1995. Improving connectionist energy minimization. *J. Artif. Int. Res.* 3(1):223–248.