T-301-REIR, Reiknirit
Fall 2019
D7 - Compression and Satisfiability

Reading material on Satisfiability (relating to questions 1-4) is in special notes, see Modules in Canvas). Material on NP-completeness (questions 6-7) is pp.903–921 in SedgewickWayne.

1. Consider the CNF formula

$$x_1x_2, \overline{x_1x_2}, \overline{x_1x_3}, \overline{x_2x_3}, x_2x_3$$

a) Find a satisfying assignment for the formula, and

That would be: x_1 : F x_2 : T x_3 : F

Proof:

x_1 or x_2 → 0 or 1 → F or T → T

not(x_1) or not(x_2) → 1 or 0 → T or F → T

not(x_1) or not(x_3) → 1 or 1 → T or T → T

not(x_2) or not(x_3) → 0 or 1 → F or T → T

x_2 or x_3 → 1 or 0 → T or F → T

which gives us then T = True

b) Give a falsifying 3-clause (i.e., a clause with 3 literals that, if added to the formula, makes the formula false)

For example: x_1 : T x_2 : T x_3 : T

Then:

x_1 or x_2 → 1 or 1 → T or T → T

not(x_1) or not(x_2) → 0 or 0 → F or F → F

which gives us a contradiction and the result is therefoere F = False

2. (8%) Perform unit propagation on the CNF formula

$$x_1x_2, \overline{x_2}, x_2x_4, \overline{x_1}x_2\overline{x_4},$$

We identify the unit literal x_1 : T, which leads to the unit literal x_2 : F.

The formula now simplifies to: (x_1 : T, x_2 : F) = x_4, not(x_4)

As the remaining unit clauses are x_4 and not(x_4), which cannot be simutlaneously be true, we derive that the CNF formula unsatisfiable.

3. Perform pure literal elimination on CNF formula

$$x_1 x_2 \overline{x_3},\ \overline{x_2 x_3},\ x_2 \overline{x_3} x_4,\ \overline{x_1} x_2 \overline{x_4},\ x_1 x_3 x_5\ .$$

- If **<u>x_1 : T</u>**, we can take out every variable of x_1 from the other parts of the CNF formula:

  x_2(not)x_3, (not)x_2(not)x_3, x_2(not)x_3x_4, x_2(not)x_4, x_3x_5

- If then **<u>x_2 : T</u>**, we can take out every variable of x2 from the other parts:

  (not)x_3, (not)x_3, (not)x_3x_4, (not)x_4, x_3x_5

- Resulting that we have a single variable of **<u>(not)x_3 : T</u>**, given us the formula:

  x_4, (not)x_4, x_5

- As can be seen be the rest of the remaining formula, we have that **<u>x_5 : T</u>**, but we also have a contradiction with <u>variable x_4, as it has to be x_4 : T and also (not)x_4 : T</u>.

  Thereby, we do not have a feasible way for this particular CNF formula.

4. (20%) The following CNF formula was treated in class (Tue 30 Oct, slide 21):

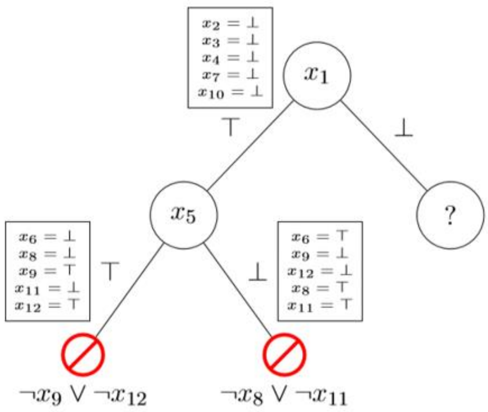
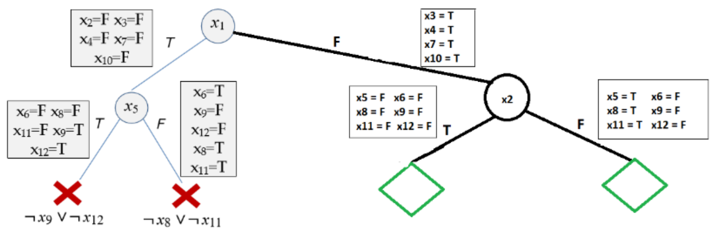


Figure 1: Partial run of the DPLL heuristic

$$\begin{aligned}\mathcal{F}':\quad & x_1x_2x_3,\ \overline{x_1}\,\overline{x_2},\ \overline{x_1}\,\overline{x_3}, \overline{x_2}\,\overline{x_3},\ x_4x_5x_6,\ \overline{x_4}\,\overline{x_5},\ \overline{x_4}\,\overline{x_6},\ \overline{x_5}\,\overline{x_6},\\ & x_7x_8x_9,\ \overline{x_7}\,\overline{x_8},\ \overline{x_7}\,\overline{x_9},\ \overline{x_8}\,\overline{x_9},\ x_{10}x_{11}x_{12},\ \overline{x_{10}}\,\overline{x_{11}},\ \overline{x_{10}}\,\overline{x_{12}},\ \overline{x_{11}}\,\overline{x_{12}},\\ & \overline{x_1}\,\overline{x_4},\ \overline{x_2}\,\overline{x_5},\ \overline{x_3}\,\overline{x_6},\ \overline{x_1}\,\overline{x_7},\ \overline{x_2}\,\overline{x_8},\ \overline{x_3}\,\overline{x_9},\ \overline{x_1}\,\overline{x_{10}},\ \overline{x_2}\,\overline{x_{11}},\ \overline{x_3}\,\overline{x_{12}},\\ & \overline{x_4}\,\overline{x_7},\ \overline{x_5}\,\overline{x_8},\ \overline{x_6}\,\overline{x_9},\ \overline{x_4}\,\overline{x_{10}},\ \overline{x_5}\,\overline{x_{11}},\ \overline{x_6}\,\overline{x_{12}},\ \overline{x_7}\,\overline{x_{10}},\ \overline{x_8}\,\overline{x_{11}},\ \overline{x_9}\,\overline{x_{12}}\ .\end{aligned}$$

The result of the initial trace of $\hat{\mathcal{F}}$ that was done in class (and lecture slides) is shown in Fig. 1. Continue the trace for the subtree given by the partial assignment x_1 = F. Label the edges with the unit clauses that were propagated. On "dead-end" states (shown on slide with a red X), identify the clause that is responsible. Assume that next variable explored is always the lowest unassigned one.

- Answer:



5. Give the LZW encoding the string aN consisting of N repeats of the character a. What is the compression ratio as a function of N?

- The LZW encoding will store and use the codewords a, aa, aaa, etc. The compression ration as a function of N would then be: O(1 / sqrt(N)) in accordance to the given information with the assignment (but I must admit I don't know why or how).

6. (Problem 5.5.18) Let F_k be the k-th Fibonacci number. Consider N symbols, where the k-th symbol has frequency F_k. Note that $F_1 + F_2 + ...F_N = F_{N+2} - 1$. Describe the Huffman code. (Hint: The longest codeword has length N −1).

- Huffman coding is a particular type of optimal prefix code that is commonly used for lossless data compression. The output from Huffman's algorithm can be viewed as a variable length code table for encoding a source symbol. The algorithm derives this table from the estimated probability or frequency of occurrences, thereby arranging the trie with the most common symbols higher up the tree, than those that are fairly uncommon (making them be at the bottom of the tree, represented by longer bit-strings). This particular tree would therefore be a chain, where the length of the i-th codeword would be i.

7. (10%) (Problem 6.62.) Suppose that P != NP. Which of the following can we infer:

e. If X is NP-complete, then X cannot be solved in polynomial time.

f. If X is in NP, then X cannot be solved in polynomial time.

g. If X is in NP but not NP-complete, then X can be solved in polynomial time.

h. If X is in NP, then X is not NP-complete.

8. (16%) Classify the following problems as in P, as NP-complete, or neither. Identify the usual name for the (underlying abstract) problem.

(a) Given a graph, is it possible to mark each node by "X", "Y" or "Z", such that neighbors always get a different mark?

(b) Given a set X of numbers $x_1, x_2, \ldots, x_n$, are there two sets X_1, X_2 such that $X_1 \cup X_2 = X$ and $X_1 \cap X_2 = \emptyset$ and such that $\sum_{x_i \in X_1} x_i = \sum_{x_i \in X_2} x_i$?

(c) Given an integer k and an virtual environment consisting of n avatars, where any two are either friends or enemies, are there k avatars that are all friends of each other?

(d) Given a computer network and two nodes a and b, and an integer k, is it possible to forward a message from a to b using at most k intermediate nodes?