CS 511, Fall 2018, Handout 28: the MaxSAT Problem and Bayesian Networks

Assaf Kfoury

November 12, 2018

Assaf Kfoury, CS 511, Fall 2018, Handout 28

page 1 of 20

(unweighted) MaxSAT problem

- Reminder: A literal is a propositional atom or negated propositional atom.
- Reminder: A clause is a disjunction of literals.
- Reminder: A propositional wff in CNF is a conjunction of clauses.
- **MaxSAT Problem**: Given a propositional wff φ in CNF, determine a truth-value assignment σ for the propositional atoms such that the number of clauses in φ satisfied by σ is maximized.

(unweighted) MaxSAT problem

- Reminder: A literal is a propositional atom or negated propositional atom.
- Reminder: A clause is a disjunction of literals.
- Reminder: A propositional wff in CNF is a conjunction of clauses.
- **MaxSAT Problem**: Given a propositional wff φ in CNF, determine a truth-value assignment σ for the propositional atoms such that the number of clauses in φ satisfied by σ is maximized.
- **Example**: Let $\varphi \triangleq (x_0 \lor x_1) \land (x_0 \lor \neg x_1) \land (\neg x_0 \lor x_1) \land (\neg x_0 \lor \neg x_1)$

 φ is **not** satisfied by any truth-value assignment σ (check it out!).

However, it is possible to assign truth-values to $\{x_0, x_1\}$ so that three out of four clauses are *true*.

Hence, given φ above as an instance of **MaxSAT** (not **SAT**), one solution is the assignment σ s.t. $\sigma(x_0) = true$ and $\sigma(x_1) = false$ which satisfies three clauses in φ . There are other solutions.

MaxSAT problem: weighted CNF formulas

- A weighted clause is a clause together with a weight, a weight being always a positive number.
- A weighted wff in CNF is a conjunction of weighted clauses.
- (Weighted) MaxSAT Problem: Given a weighted wff φ in CNF, determine a truth-value assignment σ that maximizes the sum of the weights of the clauses satisfied by σ in φ .

MaxSAT problem: weighted CNF formulas

- A weighted clause is a clause together with a weight, a weight being always a positive number.
- A weighted wff in CNF is a conjunction of weighted clauses.
- (Weighted) MaxSAT Problem: Given a weighted wff φ in CNF, determine a truth-value assignment σ that maximizes the sum of the weights of the clauses satisfied by σ in φ .

• Example:
$$\varphi \triangleq (x_0 \lor x_1)^{-5} \land (x_0 \lor \neg x_1)^{-6} \land (\neg x_0 \lor x_1)^{-2} \land (\neg x_0 \lor \neg x_1)^{-4}$$

For every assignment $\sigma : \{x_0, x_1\} \rightarrow \{true, false\}$, the following table determines the value of $\sigma(\varphi)$:

$\sigma(x_0)$	$\sigma(x_1)$	$\sigma(\varphi)$
true	true	5+6+.2=11.2
true	false	5 + 6 + .4 = 11.4
false	true	5 + .2 + .4 = 5.6
false	false	6 + .2 + .4 = 6.6

A solution (the only one here) is: $\sigma(x_0) = true$ and $\sigma(x_1) = false$.

Basic Notions and Notations of Probability Theory

- Given a *domain* or *sample space* \mathcal{D} , an *event* is a subset of \mathcal{D} .
- A probability function *P* on a finite sample space \mathcal{D} assigns to each event $A \subseteq \mathcal{D}$ a number P(A) in the closed interval of real numbers [0, 1] such that:
 - (i) P(D) = 1, and
 - (ii) $P(A \cup B) = P(A) + P(B) P(A \cap B)$.

The number P(A) is called the *probability that A occurs*.

• Computing the probability of an event *A*, given that an event *C* occurs, means finding which fraction of the probability of *C* is also in the event *A*.

This is called the *conditional probability of A given C*, and is written P(A|C).

If
$$P(C) > 0$$
, then $P(A|C) = \frac{P(A+C)}{P(C)}$ and $P(A|C) \cdot P(C) = P(A \cap C)$.

▶ Instead of $P(A \cap C)$, we write P(A, C) more compactly.

Chain rule, as a product of conditional probabilities:

$$P(A_1, \dots, A_n) = P(A_n | A_1, \dots, A_{n-1}) \cdot P(A_1, \dots, A_{n-1})$$

= $P(A_n | A_1, \dots, A_{n-1}) \cdot P(A_{n-1} | A_1, \dots, A_{n-2}) \cdot P(A_1, \dots, A_{n-2})$
= $\dots = \prod_{i=1}^n P(A_i | A_1, \dots, A_{i-1})$

Bayesian networks: An Example

A Bayesian network is a pair $(\mathcal{G}, \mathcal{P})$ where $\mathcal{G} = (V, E)$ is a DAG, such as:



Bayesian networks: An Example

A Bayesian network is a pair $(\mathcal{G}, \mathcal{P})$ where $\mathcal{G} = (V, E)$ is a DAG, such as:



- P is a set of CPT's (next defined ▶ here), with one CPT for each random variable in {x₁,...,x₈}.
- ▶ *V* is a set of vertices, one for each random variable in $\{x_1, \ldots, x_8\}$.
- ► *E* is a set of edges, describing the dependencies between variables, *e.g.*:
 - each of $\{x_1, x_2, x_3\}$ is independent of any other variable, for which we write $\pi(x_1) = \pi(x_2) = \pi(x_3) = \emptyset$ and say they have no parents.
 - ► x_4 depends on $\{x_1, x_2\}$ for which we write $\pi(x_4) = \{x_1, x_2\}$ and say the parents of x_4 are $\{x_1, x_2\}$.
 - ► x_5 depends on $\{x_2, x_3\}$ for which we write $\pi(x_5) = \{x_2, x_3\}$ and say the parents of x_5 are $\{x_2, x_3\}$, etc.

Bayesian networks: What is a CPT?

- Assume we have a Bayesian network $(\mathcal{G}, \mathcal{P})$, where $\mathcal{G} = (V, E)$ is a DAG.
- ► Every $x \in V$ can be assigned a value from a **finite domain** \mathcal{D}_x , the sample space of *x*, which is not necessarily the same for all the variables.
- ▶ Let $W \subseteq V$ be a subset of variables. An **instantiation** p of W maps every $x \in W$ to a value in \mathcal{D}_x . We can write $\{x \mapsto p(x) | x \in W\}$ for p.
- ► Let $W_1, W_2 \subseteq V$. An instantiation of W_1 and an instantiation of W_2 are **compatible** if they map every $x \in W_1 \cap W_2$ to the same value in \mathcal{D}_x .
- Every $x \in V$ is associated with a **conditional probability table** (CPT).

The CPT for variable $x \in V$ is a function T that maps each instantiation of $(\{x\} \cup \pi(x))$ to a probability in [0, 1] s.t. for every instantiation p of $\pi(x)$:

$$\sum_{u\in\mathcal{D}_x} T(\{x\mapsto u\}\cup \boldsymbol{p}) = 1$$

Exercise: Explain the reason for this restriction on the function *T*.

Bayesian networks: What is an MPE?

- An instantiation I of a subset W ⊆ V is called an evidence (also called a list of observations in Homework Assignment #9).
- Given an evidence \mathcal{I} , a **most probable explanation** (MPE) for \mathcal{I} is an instantiation \mathcal{I}' of all the variables in V with the highest probability such that \mathcal{I} and \mathcal{I}' are compatible.

MPE Problem: Given a Bayesian network $(\mathcal{G}, \mathcal{P})$ together with an evidence \mathcal{I} , find an MPE \mathcal{I}' (not necessarily unique) for \mathcal{I} .

For simplicity here, assume all the variables in $\{x_1, \ldots, x_8\}$ range over $\{a, \overline{a}\}$, *i.e.*:

$$\mathcal{D}_{x_1} = \mathcal{D}_{x_2} = \mathcal{D}_{x_3} = \mathcal{D}_{x_4} = \mathcal{D}_{x_5} = \mathcal{D}_{x_6} = \mathcal{D}_{x_7} = \mathcal{D}_{x_8} = \{a, \overline{a}\}$$

The sample space is $\mathcal{D} \triangleq \mathcal{D}_{x_1} \times \cdots \times \mathcal{D}_{x_8}$ which contains a total of $2^8 = 256$ instantiations of the random variables $\{x_1, \ldots, x_8\}$.

- To simplify the notation, it is often useful to list the variables in a fixed order, but otherwise arbitrary. We here choose the ordering *x* ≜ (*x*₁, *x*₂, ..., *x*₈), so that:
 - ► If *A* is the event of all instantiations of *x* s.t. $x_1 = x_2 = \cdots = x_5 = a$, then: $A = \left\{ (\underbrace{a, a, a, a, a}_{1, u_1, u_2, u_3}) | u_1, u_2, u_3 \in \{a, \overline{a}\} \right\}$

► If *B* is the event of all instantiations of *x* such that $x_7 = x_8 = \overline{a}$, then: $B = \left\{ (v_1, v_2, v_3, v_4, v_5, v_6, \overline{a}, \overline{a}) | v_1, v_2, v_3, v_4, v_5, v_6 \in \{a, \overline{a}\} \right\}$

•
$$A \cap B = \left\{ (\underbrace{a, a, a, a, a}_{u, \overline{a}, \overline{a}}) | u \in \{a, \overline{a}\} \right\}$$

• If C is the event that instantiates all the variables to the value a, then: $C = \{(\underbrace{a, a, a, a, a, a, a, a, a}_{\$})\}$

Make sure you understand the calculations below, which all use the chain rule (defined here) and the DAG-defined dependencies (here):

For events A, B, $(A \cap B)$, and (B|A) we can compute their probabilities as follows:

$$P(A) = P(x_1 = a) \cdot P(x_2 = a) \cdot P(x_3 = a) \cdot P(x_4 = a | x_1 = x_2 = a) \cdot P(x_5 = a | x_2 = x_3 = a)$$

$$P(B) = \sum_{(v_1, \dots, v_6, \bar{a}, \bar{a}) \in \mathcal{D}} P(x_1 = v_1) \cdot P(x_2 = v_2) \cdot P(x_3 = v_3) \cdot P(x_4 = v_4 | x_1 = v_1, x_2 = v_2) \cdot P(x_5 = v_5 | x_2 = v_2, x_3 = v_3) \cdot P(x_6 = v_6 | x_5 = v_5) \cdot P(x_7 = \bar{a} | x_4 = v_4) \cdot P(x_8 = \bar{a} | x_4 = v_4, x_5 = v_5)$$

$$P(A \cap B) = P(A) \cdot \left(\underbrace{P(x_6 = a | x_5 = a)}_{=1} + P(x_6 = \bar{a} | x_5 = a)}_{=1} \right) \cdot \frac{P(x_7 = \bar{a} | x_4 = a)}_{=1} \cdot P(x_8 = \bar{a} | x_4 = x_5 = a)}_{=1}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = P(x_7 = \bar{a} | x_4 = a) \cdot P(x_8 = \bar{a} | x_4 = x_5 = a)$$

For event *C*, which is a subset of event *A*, we can write:

 $P(C) = P(A) \cdot P(x_6 = a | x_5 = a) \cdot P(x_7 = a | x_4 = a) \cdot P(x_8 = a | x_4 = x_5 = a)$

Let I be the instantiation of the subset {x₅, x₆, x₇, x₈} such that I(x₅) = I(x₆) = I(x₇) = I(x₈) = ā. Let D be the set of all instantiations I' of the full set {x₁,...,x₈} that are compatible with I. Hence, D is an event containing 2⁴ instantiations I' (why?), each of the form:

$$\mathcal{I}' = (v_1, v_2, v_3, v_4, \overline{a}, \overline{a}, \overline{a}, \overline{a})$$

where $v_1, v_2, v_3, v_4 \in \{a, \overline{a}\}$ and whose probability is:

$$P(\mathcal{I}') = P(x_1 = v_1) \cdot P(x_2 = v_2) \cdot P(x_3 = v_3) \cdot P(x_4 = v_4 | x_1 = v_1, x_2 = v_2) \cdot P(x_5 = \overline{a} | x_2 = v_2, x_3 = v_3) \cdot P(x_6 = \overline{a} | x_5 = v_5) \cdot P(x_7 = \overline{a} | x_4 = v_4) \cdot P(x_8 = \overline{a} | x_4 = v_4, x_5 = v_5)$$

Exercise: Using the CPT's defined here, find an instantiation $\mathcal{I}' \in D$, compatible with the instantiation \mathcal{I} such that $\mathcal{I}(x_5) = \mathcal{I}(x_6) = \mathcal{I}(x_7) = \mathcal{I}(x_8) = \overline{a}$, for which the probability $P(\mathcal{I}')$ is maximized.

Hint: You can solve the problem by hand. No need to use a computer. (This exercise solves the MPE Problem for a particular relatively easy case.)

Let T_i be the CPT of random variable x_i in tabular form, where $1 \le i \le 8$.

$$T_{1} \begin{cases} \frac{x_{1}}{a} & \frac{P(x_{1})}{a} & .3 \\ \overline{a} & .7 \end{cases} T_{2} \begin{cases} \frac{x_{2}}{a} & \frac{P(x_{2})}{a} & .2 \\ \overline{a} & .2 \\ \overline{a} & .3 \end{cases} T_{3} \begin{cases} \frac{x_{3}}{a} & \frac{P(x_{3})}{a} & .9 \\ \overline{a} & .9 \\ \overline{a} & .1 \end{cases}$$

$$T_{4} \begin{cases} \frac{x_{1}}{a} & x_{2} & x_{4} & \frac{P(x_{4}|\{x_{1}, x_{2}\})}{a} & .1 \\ \frac{a}{a} & a & \overline{a} & .2 \\ a & a & \overline{a} & .2 \\ a & a & \overline{a} & .8 \\ a & \overline{a} & a & .1 \\ \overline{a} & a & a & .1 \\ \overline{a} & \overline{a} & \overline{a} & .9 \\ \overline{a} & \overline{a} & \overline{a} & .6 \\ \overline{a} & \overline{a} & \overline{a} & .7 \\ \overline{a} & \overline{a} & \overline{a} & .1 \\ \overline{a} & \overline{a} & \overline{a} & .9 \\ \overline{a} & \overline{a} & \overline{a} & .1 \\ \overline{a} & \overline{a} & \overline{a} & .1 \\ \overline{a} & \overline{a} & \overline{a} & .9 \\ \overline{a} & .1 \\ \overline{a} & \overline{a} & \overline{a} & .9 \\ \overline{a} & .1 \\ \overline{a} & \overline{a} & \overline{a} & .9 \\ \overline{a} & .1 \\ \overline{a} & \overline{a} & \overline{a} & .9 \\ \overline{a} & .1 \\ \overline{a} & \overline{a} & \overline{a} & .9 \\ \overline{a} & .1 \\ \overline{a} & \overline{a} & \overline{a} & .9 \\ \overline{a} & .1 \\ \overline{a} & \overline{a} & \overline{a} & .9 \\ \overline{a} & .1 \\ \overline{a} & \overline{a} & \overline{a} & .9 \\ \overline{a} & .1 \\ \overline{a} & \overline{a} & \overline{a} & .9 \\ \overline{a} & .1 \\ \overline{a} & \overline{a} & \overline{a} & .9 \\ \overline{a} & .1 \\ \overline{a} & \overline{a} & \overline{a} & .9 \\ \overline{a} & .1 \\ \overline{a} & \overline{a} & \overline{a} & .9 \\ \overline{a} & .1 \\ \overline{a} & .1 \\ \overline{a} & \overline{a} & \overline{a} & .9 \\ \overline{a} & .1 \\ \overline$$

								1	<i>x</i> ₄	<i>x</i> 5	<i>x</i> ₈	$P(x_8 \{x_4, x_5\})$
									a	а	а	.4
	x5	x_6	$P(x_6 \{x_5\})$		$\int x_4$	<i>x</i> ₇	$P(x_7 \{x_4\})$		а	а	\overline{a}	.6
	а	а	.1		а	а	.7		а	\overline{a}	а	.5
T_6	а	\overline{a}	.9	T_7	а	\overline{a}	.3	T_8	а	\overline{a}	\overline{a}	.5
	ā	а	.5		ā	а	.4		ā	а	а	.6
	L ā	\overline{a}	.5		L ā	\overline{a}	.6		\overline{a}	а	\overline{a}	.4
									ā	\overline{a}	а	.7
								l l	ā	\overline{a}	\overline{a}	.3

Reducing MPE to MaxSAT

Consider the CPT T_i of random variable x_i . In tabular form, T_i is a set of n_i rows, denoted $r_{i,1}, r_{i,1}, \ldots, r_{i,n_i}$, and we can thus write:

$$T_i = \{r_{i,1}, r_{i,1}, \ldots, r_{i,n_i}\}$$

For example, for CPT T_6 defined \checkmark here which has $n_6 = 4$ rows, we can write:

$$T_{6} = \left\{ [(x_{5}, a), (x_{6}, a), .1], [(x_{5}, a), (x_{6}, \overline{a}), .9], \\ [(x_{5}, \overline{a}), (x_{6}, a), .5], [(x_{5}, \overline{a}), (x_{6}, \overline{a}), .5] \right\}$$

► For the reduction from MPE to MaxSAT, we represent the pairs (x_i, a) and (x_i, \overline{a}) by propositional atom p_i and its negation $\neg p_i$: $p_i \triangleq (x_i, a)$ and $\neg p_i \triangleq (x_i, \overline{a})$. For example, substituting the representations of (x_i, a) and (x_i, \overline{a}) in T_6 produces:

$$T_{6}' = \left\{ [p_{5}, p_{6}, .1], [p_{5}, \neg p_{6}, .9], [\neg p_{5}, p_{6}, .5], [\neg p_{5}, \neg p_{6}, .5] \right\}$$

Reducing MPE to MaxSAT (continued)

Transforming a Bayesian network into a weighted CNF:

Every row r in every CPT T is transformed into a weighted clause which contains the negations of all the propositional atoms in r and is weighted with the negative logarithm of the conditional probability in r.

For example, CPT T'_6 as a set of four rows (defined rectarrow here) is transformed into:

weighted-clause $([p_5, p_6, .1]) \triangleq (\neg p_5 \lor \neg p_6)^{-\log .1}$ weighted-clause $([p_5, \neg p_6, .9]) \triangleq (\neg p_5 \lor p_6)^{-\log .9}$ weighted-clause $([\neg p_5, p_6, .5]) \triangleq (p_5 \lor \neg p_6)^{-\log .5}$ weighted-clause $([\neg p_5, \neg p_6, .5]) \triangleq (p_5 \lor p_6)^{-\log .5}$

Note: a probability *p* is always a number in the closed interval [0, 1], so that $\log p$ is a negative number and $-\log p$ is a positive number.

We obtain a weighted CNF by collecting the weighted clauses of all the rows in all the CPT's.

For our running example with 8 random variables and the DAG shown here:

induced-CNF
$$(T_1, \ldots, T_8) riangleq igwedge_{1 \leqslant i \leqslant 8, \ 1 \leqslant j \leqslant n_i}$$
 weighted-clause $(r_{i,j})$

Reducing MPE to MaxSAT (continued)

► Theorem: For any instantiation I of a Bayesian network which contains only binary variables, the sum of the weights of the clauses that I leaves unsatisfied in the induced CNF is equal to - log P(I).

This is Theorem 1 in *Using Weighted MAX-SAT Engines to Solve MPE* by J. D. Park, which also includes its proof.

Corollary: Maximizing the weight of the satisfied clauses minimizes the sum of the unsatisfied clauses, which is equivalent to maximizing the probability in the original Bayesian network. Thus, solving the MaxSAT problem also solves the MPE problem.

Reducing MPE to MaxSAT (continued)

- The reduction from the MPE problem to the MaxSAT problem as described in the preceding pages is limited to the case when:
 - 1. All the probabilities in the CPT's are non-zero, and
 - 2. All the random variables are binary.
- The article by J.D. Park explains how to extend the method in order to lift both limitations (how to handle zero probabilities and how to allow non-binary random variables).

(THIS PAGE INTENTIONALLY LEFT BLANK)